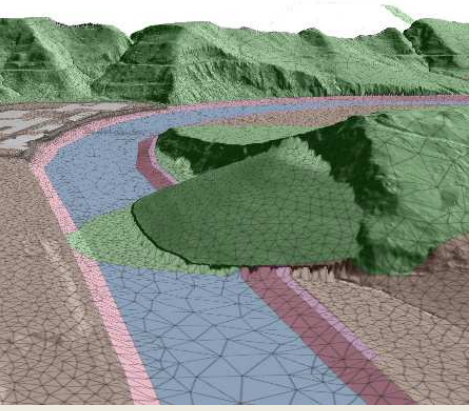
Extremhochwasser an der Aare. Detailbericht A Projekt EXAR. Hydrometeorologische Grundlagen

Staudinger, Maria ; Viviroli, Daniel



Extremhochwasser an der Aare

Detailbericht A Projekt EXAR

Hydrometeorologische Grundlagen

Fotos Umschlag

1. Stauanlage Mühleberg, Mühleberg. Foto: IUB Engineering AG, Begehung vom 5. April 2017.
2. Mündungsbereich Hagneckkanal in Bielersee. Foto: IUB Engineering AG, Begehung vom 10. April 2017.
3. Höhenmodell der Rutschung Brättele, nahe des Beurteilungsperimeters Mühleberg. Quelle: GEOTEST AG, 2020.
4. Wasser- und Kernkraftwerk bei Beznau. Foto: Schweizer Luftwaffe 2005, © VBZ.
5. Verklausung der Mattenschwelle bei Bern während des Hochwassers im August 2005. Foto: Schweizer Luftwaffe 2005, © VBZ.

Impressum
Datum: 06.11.2020

Herausgeber
Universität Zürich, Geographisches Institut, Winterthurerstrasse 190, CH-8057 Zürich

Auftraggeber
Bundesamt für Umwelt (BAFU), Eidgenössisches Nuklearsicherheitsinspektorat (ENSI), Bundesamt für Energie (BFE), Bundesamt für Meteorologie und Klimatologie (MeteoSchweiz), Bundesamt für Bevölkerungsschutz (BABS)

Projektleitung
Daniel Viviroli, Jan Seibert
Universität Zürich

Experten/innen
Helmut Habersack, BOKU Wien
Douglas Maraun, Universität Graz
Christian Pfister, em. Universität Bern
Bettina Schaefli, Universität Lausanne, Universität Bern
Anton Schleiss, em. École polytechnique fédérale de Lausanne EPFL
Bruno Sudret, ETH Zürich

Redaktion
Maria Staudinger, Daniel Viviroli
Universität Zürich

Autoren und Mitarbeit Inhalte Detailbericht

Universität Zürich	Jan Seibert, Anna E. Sikorska-Senoner, Maria Staudinger, Daniel Viviroli, Ida Westerberg
Universität Bern	Martina Kauzlaric, Luise Keller, Rolf Weingartner
Grenoble INP	Jérémy Chardon, Guillaume Evin, Anne-Catherine Favre, Benoit Hingray, Gilles Nicolet, Damien Raynaud
EPFL	Peiman Asadi, Anthony Davison, Sebastian Engelke, Thomas Lugrin
PSI	Vinh Dang, Calvin Whealton
HZP	Matthias Pfäffli

Zitiervorschlag
Staudinger, M., Viviroli, D. 2020: Extremhochwasser an der Aare. Detailbericht A Projekt EXAR. Hydrometeorologische Grundlagen. Universität Zürich. Zürich: 383 S.

Hinweis
Dieser Bericht wurde im Auftrag der Bundesämter BAFU, ENSI, BFE, MeteoSchweiz und BABS verfasst. Für den Inhalt ist allein der Auftragnehmer verantwortlich.

Contents

1 Introduction

Daniel Viviroli

The hydro-meteorological simulations elaborated by EXAR's AP2 are an important basis for all following steps taken and assessments made in the project. Due to this is importance, the simulations need to be provided with comprehensive background, in order to assess their plausibility and to reveal relevant meteorological and hydrological patterns. Such background is especially important for large floods, as the simulated discharge peaks of very rare and extremely rare events (probability of less than 10^{-4} per year) are higher than could be expected from statistical analyses of observed discharge peaks. In particular, the increase of peak discharge with decreasing probability does not level off for very rare and extremely rare events when plotted as an exceedance curve. This behaviour is neither implausible nor surprising, but it still calls for a thorough analysis to strengthen confidence in the results. It has to be mentioned that an assessment of plausibility is challenging due to the lack of sufficiently long records of observational data. Therefore, analyses in this direction rather focus on plausibility of patterns and sensitivities of the model components involved.

In this sense, Chapter 2 contains further background on the GWEX weather generator, introduces an alternative weather generator SCAMP along with similar background, and discusses the generation of large precipitation events. Chapter 3 contains sensitivity and plausibility analyses of the hydrological simulations (HBV and RS Minerve). As noted in the main report, none of the comprehensive analyses done has revealed significant shortcomings of methods or results. The analyses show that major critical processes and patterns are represented adequately, and that the flood events simulated are a suitable basis for the risk and hazard assessments made in EXAR.

Further, this detail report contains several additional chapters that serve to discuss important details of the simulations or, regarding the appendices, provide full sets of figures that could not be placed in the main report:

- Chapter 4 provides background on an inconsistency in the hydrological simulations and how this was amended.
- Chapter 5 discusses the possibility of an overflow of the Rhine River into the Aare River in the region of Mels.
- Chapter 6 shows details about rating curve uncertainty for all sites where such analyses were feasible.
- Chapter 7 describes in more detail how the Jura Lakes and the weir at Port were implemented in RS Minerve
- Chapter 8 finally contains the exceedance curves for all transfer points, where available with information on observations, extrapolated observations, and historical floods.

2 Weather Generators: GWEX and SCAMP

Jérémy Chardon, Guillaume Evin, Anne-Catherine Favre, Benoit Hingray, Gilles Nicolet, Damien Raynaud

Two distinctly different weather generators have been developed to provide very long time series of hourly precipitation and temperature for multiple sites. GWEX is a multisite, multivariate two-part stochastic weather generator (as introduced in the main report) and the alternative weather generator SCAMP is a hybrid weather generator based on atmospheric and weather analogs. Both weather generators were applied to create precipitation and temperature time series at hourly resolution for 105 sites and for very long time periods (30 x 10'000 years each). Note that a number of different parameterisations havexcee been developed and compared for each weather generator. In the following, we only describe the ones finally retained for the generation of the weather scenarios further used by AP2 and AP3 partners within EXAR.

2.1 GWEX weather generator

GWEX is a multi-site precipitation and temperature model aiming at reproducing the statistical behaviour of weather events, with a focus on extremes, at different temporal (*n*-day amounts) and spatial scales (e.g. at the stations and for different sub-catchments). It relies strongly on the structure proposed by Wilks (1998), for which daily precipitation amounts $Y_t(k)$ are generated independently of daily precipitation occurrences X_t(k), with several major improvements (Evin *et al.*, 2018). To improve the simulation of 3-day precipitation amounts, GWEX is applied first to 3 days precipitation amounts and then disaggregated to hourly scenarios using meteorological analogs. Here, we further describe the first module of GWEX, related to the generation of 3-days multisite precipitation amounts. Details on the model and especially on the disaggregation step are described in Evin *et al.* (2018) and Appendix 10.3.1.

2.1.1 Precipitation occurrence process

As proposed by Wilks (1998), the occurrence process is represented for each station by a two-state first order Markov chain (as 3-day precipitation amounts are considered, the temporal dependence is taken into account over 6 days), representing 'dry' and 'wet' states (a state referring to a 3-days sequence) so that:

$$X_t(k) = \begin{cases} 0, & \text{if dry conditions at location k at time step } t. \\ 1, & \text{if wet conditions at location k at time step } t. \end{cases}$$

For any given rainfall station, a dry/wet state corresponds to a precipitation amount smaller/greater than a threshold equal to 0.5 mm. The seasonality of the occurrence process is taken into account by estimating model parameters at a monthly scale.

Parameter Estimation. At each station, the probability of a wet state at time step *t* is given by the transition probability directly estimated by the proportion of wet time steps (e.g.; time steps for which $X_t(k) = 1$) following the current observed wet or dry state ($X_{t-1}(k)$).

Inter-site correlation. Following Wilks (1998), inter-site correlations between precipitation occurrences are introduced using a multivariate Gaussian distribution. For two sites *k* and *l*, a bivariate normal distribution with mean **0**, variance **1** and a correlation parameter $\omega(X_t(k), X_t(l))$ is used to

reproduce the empirical (i.e. observed) correlation $\hat{\rho}(X_t(k), X_t(l))$. The relationship between ω and $\hat{\rho}$ is not direct since the at-site occurrence process also influences $\hat{\rho}$ (Wilks, 1998). In practice, ω is found iteratively, stochastic simulations being generated for different values of ω and given transition probabilities, until the evaluation of the correlation between the simulated precipitation occurrences ρ matches $\widehat{\rho}$. The cross-correlations, ω, are estimated for all possible pairs of stations using the non-parametric Kendall's correlation coefficient τ, which can be directly related to the Pearson correlation coefficient ρ (McNeil *et al.*, 2005, p.97).

2.1.2 Precipitation intensity process

Similar to the occurrence process, the seasonal aspect of the precipitation intensity is taken into account by fitting the model for each month, accounting for a 3-months moving window for each month.

For each station within the region, the distribution used to model and generate 3 days precipitation amounts is the Extended Generalized Pareto-Type III Distribution (E-GPD, Papastathopoulos and Tawn, 2013) which has been shown to model precipitation intensities adequately (Naveau *et al.*, 2016). This distribution can be described by a smooth transition between a gamma-like distribution and a heavy-tail Generalized Pareto distribution (GPD). The transition is obtained via a transformation function $G(v) = v^{\kappa}$ such that the whole range of precipitation intensities is modelled without a threshold selection (Naveau *et al.*, 2016).

Parameter estimation. The shape parameter of the E-GPD distribution (ξ) was estimated using a robust advanced regionalization method (Evin *et al.*, 2016). Bootstrap analyses were carried to demonstrate the fairness of this procedure (see Appendix 10.3.5). For each station, the method introduced all precipitation data available from the neighbouring stations with a similar statistical behaviour of their maximum values. The shape parameter is also estimated on a seasonal basis. The two remaining parameters, the scale parameter σ and the parameter of the transformation κ, are estimated at each station using the method of probability weighted moments (Naveau *et al*., 2016).

Space-Time Correlations. Let Z_t denote a vector of *K* Gaussian variables with mean **0** defined as $Z_t(k) = \Phi^{-1}\left[\hat{F}\{Y_t(k)\}\right]$ where $\hat{F}$ is the empirical distribution function and Φ indicates the standard normal cumulative distribution function. Spatial and temporal dependence of precipitation amounts is represented using a Multivariate Autoregressive model of order 1, MAR(1), described as follows:

$$Z_t = AZ_{t-1} + \varepsilon_t$$

where A is a $K \times K$ matrix and ε_t is a random $K \times 1$ noise vector. The elements of ε_t have zero means and are independent of the elements of Z_{t-1}. The covariance matrix of ε_t noted Ω_Z can be expressed as $\Omega_Z = M_0 - AM_0'A$ where M_0 is the covariance matrix of Z_t and M_0' denotes its transpose. Following Bárdossy and Pegram (2009), A is taken to be a diagonal matrix, the diagonal elements being the lag-1 serial correlation coefficients of the intensity process $\mathrm{Y_t(k)}$. Direct estimates of M_0 and A cannot be obtained since non-zero precipitation amounts $Y_t(k)$ are not observed. Following the methodology proposed by Wilks (1998) and Keller *et al.* (2015), for each pair of stations (k, l), long sequences of precipitation amounts were generated using the estimated parameters of the occurrence process. Each element $\widehat{m}_0(k, l)$ of the matrix M was then found iteratively by matching the correlation between these long random streams with the observed correlation. The estimates of the coefficient of the matrix A were obtained using the same simulation approach (Evin *et al.*, 2018).

Innovations ε_t are often assumed to follow a standard multivariate normal distribution. This is equivalent to assume that extreme precipitation realizations generated for the different sites are spatially independent, which is not realistic. In GWEX, the spatial dependence structure of ε_t is modelled with a Student copula. The additional copula parameter ν (describing the tail dependence) is estimated by likelihood-maximization, as described in McNeil *et al.* (2005, Section 5.5.3.). The Fisher copula (Favre *et al.*, 2018, Appendix 10.3.3) was also evaluated to represent the spatial dependence (it allows to account for an asymmetry between lower and upper tail dependence which can be observed for precipitation data). Since the Fisher copula did not improve simulations (Appendix 10.3.4) it was not further considered. Note that neither Student copula nor Fisher copula allow asymptotic independence (Coles *et al.*, 2000; Davison *et al.*, 2013) of extreme precipitations, i.e. the spatial dependence does not decrease with the level of extreme values.

Our analyses, carried out for all calendar months separately, suggest that precipitation extremes may have some asymptotic dependence for some months (e.g. December, January, February) and some independence for others (e.g. June, July, September). The asymptotic dependence or independence of extremes is however very difficult if not rather impossible to estimate from short times series: The 85 years of daily precipitation data available for EXAR at different station are not long enough for a robust estimation. In lack of better information, we thus considered that the spatial co-variability behaviour of extremes is the same than that estimated for all other precipitation values. This simplification is thus conservative allowing for the generation of concomitant extreme precipitation at different neighbouring sites. The approach thus allows to explore critical situations where extremes occurs at multiple sites. It is important to note that the generation process does not necessary lead to the concomitancy of extremes at different sites. It also allows the generation of localised extremes. The probability that extremes are generated at all sites at the same time is in fact almost zero. The asymptotic independence hypothesis is hard to evaluate from the available time series. For the Aare catchment, it was found to be only valid for June, July and September (see Appendix 10.3.6).

2.1.3 Multi-site temperature model

At each station, daily data of standardized temperature are modelled with a skew normal distribution; standardization is based on a calendar basis and involves both the calendar mean and the standard deviation of the current simulation day. The spatial and temporal dependence of daily temperature at multiple sites are modelled by a MAR(1) process. The sub-daily structure of temperature is finally obtained by disaggregation of daily temperatures. Temperature simulations were conditioned on precipitation. The model is described in detail in Evin et al. (2019) (paper available in Appendix 10.3.2).

2.1.4 Weather simulations

GWEX has been applied for the generation of weather scenarios at 105 precipitation stations and 26 temperature stations available within the basin (Figure 1). For both variables, 30 time series of 10'000 years have been generated and disaggregated at an hourly scale. Mean areal estimates for the 89 sub-catchments and/or for larger spatial units (mean areal precipitation (MAP) and mean areal temperature (MAT) in the following) are computed using Thiessen's polygon method.

2.2 SCAMP weather generator

The alternative weather generator, SCAMP, is made of a suite of models that aim at generating long hourly time series of MAP and MAT for each of the 89 sub-catchments considered in the HBV-model of the Aare basin.

2.2.1 Atmospheric analogs for generating long atmospheric trajectories

The first step of SCAMP is to generate long series of synoptic weather over Europe. We used for this the ERA20C atmospheric reanalysis (Poli *et al.*, 2016) available for the period 1900-2010. For the current simulation day, the *K*-nearest atmospheric analog days are identified in the 110 years covered by the ERA20C reanalyses. The analogs are identified within the days included in a moving temporal window of +/-2months centered on the simulation day (*e.g.* if the simulation day is April 1st, 1997, within all days between March, 1st and April, 30th of all 110 years). One of the *K*-nearest analogs is then randomly selected and its following day is used as large scale atmospheric scenario for the day following the current simulation day (e.g. for April 2nd, 1997). This model allows to generate very long time series of atmospheric trajectories. It allows also for the creation of new atmospheric trajectories by rearranging the atmospheric sequences observed in the last 110 years. A detailed description of the simulation process is given in Appendix 10.4.1 and in Raynaud *et al*. (2019). A 30x 10'000-year-long sequence of synoptic weather was generated with the model.

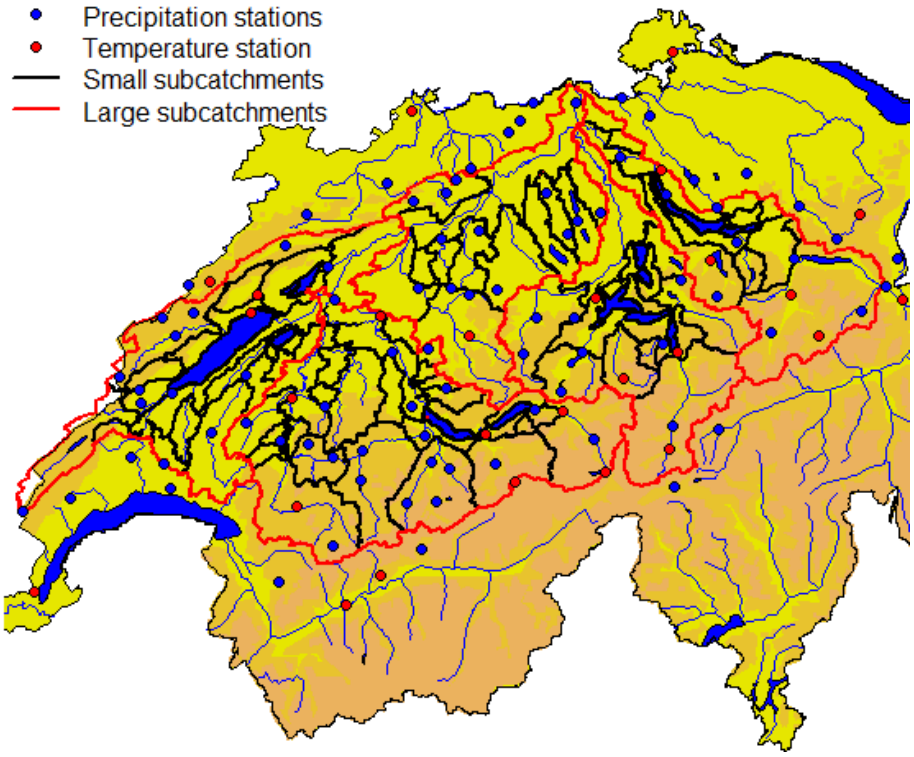


Figure 1 Aare basin with the 105 precipitation stations, the 26 temperature stations, the 89 small sub-catchments and the 5 large sub-catchments.

2.2.2 Atmospheric analogs for regional weather scenarios

In a second step, each of the 10'000-year-long time series of synoptic weather was used to generate a 10'000- year-long time series of regional daily weather, namely MAP and MAT for the Aare catchment. For this a standard model based on atmospheric analogs is used that is typically applied for downscaling purposes (Chardon *et al.* 2016, Raynaud *et al.* 2017). For each day of the 10'000-year-long time series of synoptic weather, the *K*-nearest atmospheric analogs are identified in the period 1930-2010, the period for which both large scale atmospheric reanalyses (ERA20C) and local scale weather observations (ANETZ) are available. The regional weather scenario for the day is then built from the statistical distribution of the regional weather observed for those *K*- analogs (see step 3 above). The model applied for the identification of atmospheric analogs is hierarchical: 2 different analogy levels

are considered in turn based on different analogy predictors respectively. This parameterisation was adapted from that presented in Raynaud *et al*. (2017), which guarantees both inter-variable physical consistency and good predictive skills for four predictors (precipitation, temperature, solar radiation and wind).

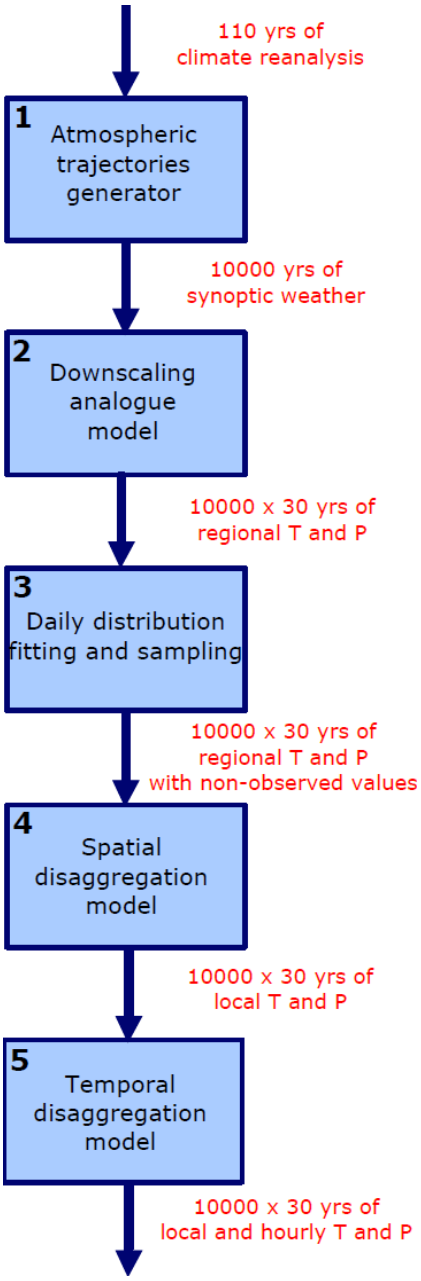


Figure 2 Steps of the SCAMP weather generator.

For the first level of analogy, large scale predictors are geopotential heights at 1000 hPa and 500 hPa (HGT1000, HGT500) (Raynaud *et al.* 2017). From June to August, the analogy is evaluated with fields available at 12UTC. From September to May, the analogy uses the geopotential fields at both the current day *D* and its following day (*D*+1) at 12UTC. This allows to better catch the motions of low pressure systems and fronts and improves the prediction skill for precipitation. The 100 nearest analogs of each simulation day are retained for the second analogy step (each day of the long time series of synoptic weather generated in previous step).

For the second level of analogy, predictors are the vertical velocities at 600 hPa and the large scale temperature at 2 m from September to May. In summer, a large part of the precipitation is due to convective storms. Vertical velocities or other more relevant predictors such as the Convective Available Potential Energy (CAPE) fail to predict convective phenomena due to the coarse resolution of the reanalysis data. Consequently, a most straightforward predictor, namely large scale precipitation, is used instead of vertical velocities. This choice has been proven to largely improve the predictive skills of the method during summer months and it allows achieving summer skill scores comparable to the ones of the rest of the year. This second analogy proceeds a sub-selection of 30 analogs within the 100 obtained from the first analogy level. The dimensions and the localisation of the different spatial windows used to evaluate the analogy between days follow the recommendations of Raynaud *et al.* (2017).

After this 2 step downscaling the analog model produces finally 30 values of daily MAP and MAT for each day of each 10'000-year-long time series of large scales atmospheric situations.

2.2.3 Distribution sampling to generate non-observed regional weather scenarios

To produce a time series of daily MAP and MAT, the standard analog method simply uses for each simulation day the MAP and MAT values observed for one of the 30 analogs identified in the previous step. Weather scenarios are then obtained by resampling observations from analog days. Consequently, it is impossible to produce daily weather time series with values out of the ranges of MAP and MAT observations. To overcome this issue, a stochastic generation process was plugged to the MAP and MAT values obtained each day with the analog model of the previous step. Practically, a Gamma (respectively Gaussian) statistical distribution is fitted to the 30 observed precipitation

(respectively temperature) values available for the 30 analog days of the current simulation day (see Chardon *et al*. 2017; for details). The fitted distribution (its parameters vary from one day to the other) is then used to generate *n* non-observed values of precipitation and temperature from random draws. In the present case, *n* is fixed to 30.

The result of the stochastic generation step is 30 non-observed values of daily MAP and MAT for each day of the simulation. Finally, for the development of the time series the temporal organisation of those values, from one time step to the other, had to be chosen. In this step MAP and MAT time series were produced independently with temporal relevance for each variable and relevant co-variability between variables. Part of the temporal/physical consistency between two consecutive days or between temperature and precipitation time series was derived from conditioning the scenarios on the large scale atmospheric situation and then on the trajectory. Additional consistency between days and variables was obtained with the Schaake Shuffle method (Clark *et al.*, 2004). This method uses both the inter-variable physical and the intra-variable temporal consistency in the observations to combine the outputs of the analog method. In practice, we applied the method hierarchically: first the temporal consistency of precipitation was forced and then temperature scenarios conditioned on obtained precipitation scenarios were generated.

2.2.4 Weather analogs for the generation of high resolution scenarios

To go from daily regional MAP and MAT scenarios to sub-daily scenarios for the 89 sub-catchments of the Aare system, a non-parametric disaggregation approach was applied, following the methodology developed by Mezghani and Hingray (2009) for the upper Rhone River. In the approach, observed spatial structures of weather analogs were identified within the 1930-2014 period. Sub-daily observations were only available for a limited set of stations and on a short time period (from 1990 to 2014). A 1930-2014 time series of pseudo-observations has been reconstructed at an hourly time step for the 89 sub-catchments and for the entire period (1930-2014) using the information from the recent period (1990-2014). The sub-daily data for all stations are obtained from the nearest meteorological analog, where the analogy is again evaluated from a meteorological point of view (based on the daily spatial patterns). The sub-daily structure for stations without sub-daily data was derived from the closest neighbouring station with sub-daily data. More details on this procedure can be found in Appendix 10.2.

For each day, the 30 MAP values issued from the distribution fitting and sampling were compared to the 30 initial MAP values of the analogs. For each MAP scenario P_{sce}, the closest analog MAP P_{ana} was identified in the 30 initial analog days. The spatial/temporal structure of this analog day was used for the disaggregation of the MAP scenario P_{sce}. Local values of the analog day observed for each time step of each of the 89 sub-catchments were then scaled with the P_{sce}/P_{ana} coefficient and further used as local scenarios for the 89 sub-catchments. Note that a set of observed reference structures were applied instead of the structure of the analog days when the P_{sce}/P_{ana} ratio of all the 30 analog days was higher than a given threshold. This safeguard prevented applying very large P_{sce}/P_{ana} scaling factors and thus generating large and unrealistic local precipitation values. The spatial structure used for temperature is the observation from the same day as the selected precipitation structure. This structure is also adapted by adding the $T_{sce} - T_{ana}$ coefficient to all local daily temperature values. More details on the functioning of the spatial disaggregation are described in Appendix 10.4.2.

2.3 Weather scenarios from GWEX and SCAMP

For each weather generator, 30 times series of high resolution (multisite/hourly) MAP and MAT scenarios were produced. Each scenario covered a time period of 10‘000 years.

In a first step, the generated time series were evaluated on their ability to reproduce a number of key characteristics and statistical behaviours of observations available for the period of 1930-2014. For precipitation, this includes the probability of wet/dry days for each month, the statistical distribution of the length of dry and wet spells, the statistical distribution of all-days and extreme precipitation amounts, the spatial correlation between sites and the dependence of those characteristics to the season. For temperature, this includes the statistical distribution of heat waves, and length of cold spells, the spatial correlations between sites, the distribution of daily maxima and minima. Most evaluations were carried out for multiple temporal and spatial scales: from daily to multiple days and monthly scales, for individual meteorological stations (105 precipitation stations, 26 temperature stations), for the 89 elementary catchments of the HBV-model of the Aare catchment, for the 5 main sub-catchments of the Aare catchment and for the whole Aare catchment. They were evaluated for each weather variable independently, as well as for derivative variables, which account for the co-variance structure between precipitation and temperature. For the latter, the capacity of each weather generator to simulate relevant correlations between precipitation and temperature was assessed with the hydrological filter of the Aare system. A rainfall-snowfall model was then applied using the observed and the simulated weather as input. The statistical characteristics of the liquid/solid precipitation time series from simulated weather inputs were then evaluated against those obtained from the observed weather (Gallou, 2018). From all our analyses and checks, the performances of GWEX and SCAMP were satisfying (see for instance Figure 4). Part of these evaluations are presented in peer reviewed publications and the reader is referred to these for further details (see Evin *et al.*, 2018, 2019 for GWEX precipitation and temperature respectively; Chardon *et al.*,2017 and Raynaud *et al.*, 2017 for SCAMP).

In the following, only the return levels of maximal precipitation for different spatial/temporal scales obtained from the 30x10’000 year time series for large return periods are presented. More results are in Appendix 10.5.1, Appendix 10.5.2 and Appendix 10.5.4. Moreover, Appendix 10.5.3 and Appendix 10.5.3 present the precipitation events that led to the ten largest simulated floods in detail.

2.4 Meteorological simulations with GWEX

As shown in Evin et al. (2018 and 2019) and mentioned above, GWEX was able to reproduce the major characteristics of precipitation and temperature observations at all spatial and temporal scales.

The empirical return levels of maximal MAP1d and MAP3d obtained from the 30 x10’000-year-long time series (according to the Gringorten plotting position formula), for the whole Aare catchment and for its 5 main sub-catchments, namely Neuchatel, Thun, Aare-Emme, Reuss and Limmat, the left side of Figure 3 and Figure 5.For the short return periods, for which return levels can be also estimated from observed MAP (with the same Gringorten formula), the return levels of the simulations are very close to the empirical ones and highlight the good performance of the model for those variables. For the whole catchment and for the Neuchatel, Thun and Aare Emme sub-catchments, most of the 18’000 years return levels were between 130 mm and 160 mm for MAP1d (respectively 190 mm and 225 mm for MAP3d). For the two most eastern sub-catchments (Reuss and Limmat), the values are slightly higher, between 160 mm and 205 mm for MAP1d (respectively 230 mm and 270 mm for MAP3d).

The bottom-left part of Figure 3 and Figure 5 maps the mean largest value for the 89 sub-catchments for MAP1d (Figure 3) and MAP3d (Figure 5), i.e. the average value of the 30 different 18'000 years return levels obtained respectively for the 30 time-series. Similar spatial patterns were found for MAP1d and MAP3d. The largest return levels are again located in the Southeast of the Aare basin (200 mm to 280 mm for MAP1d and 280 mm to 350 mm for MAP3d). Large return levels were also obtained in a sub-catchment in Canton of Fribourg, towards Jaun and La Valsainte (220 mm for MAP1d and 297 mm for MAP3d). In the West, close to the Jura Mountains, the return levels are slightly smaller (160 mm to 200 mm for MAP1d and between 210 mm and 270 mm for MAP3d). Similar results are obtained for the central part of the Aare system. The lowest return levels are situated in the North (between 120 mm and 150 mm for MAP1d and between 160 mm and 210 mm for MAP3d).

The performance of GWEX with regard to additional characteristics to extreme MAP1d and MAP3d are presented in Figure 4. Figure 4 also presents the very good performance of GWEX for the simulation of the statistical distribution of wet/dry spells and of the monthly amounts at different spatial scales (whole Aare, main sub-catchments, and stations).

2.5 Meteorological simulations with SCAMP

As shown in Chardon *et al.* (2018), Raynaud *et al.* (2018) and mentioned above, the SCAMP weather generator also allows reproducing the characteristics of precipitation and temperature observations at all spatial and temporal scales. The 30x10'000-year-long weather time series also present long-term meteorological situations that cannot be found in the observations, whatever the time scale under consideration. For all four seasons, the ranges of simulated seasonal temperature and precipitations exceed the observed ones. For instance, the minimum and maximum observed winter precipitation amounts are 60 and 490 mm respectively. In the simulations, these values reached 40 mm and 690 mm. Such characteristics are particularly interesting for hydrological purposes as they will allow simulating extreme discharge events with non-observed initial conditions in terms of soil moisture and snow pack.

The results on precipitation maxima are presented in the right side of Figure 4 and Figure 6, for the same spatial and temporal scales as before. Good agreement is obtained between observations and simulations for return periods up to 150 years (maximum return period that can be estimated with the Gringorten formula with 85 years of observed data). For the entire catchment, the 18000 years MAP1d was 140 mm on average but it reached almost 200 mm for some scenarios. These values reached 190 mm and 250 mm for MAP3d respectively, showing that for high precipitation events 75% of the total amount fell within 24h. For both MAP1d and MAP3d, the Limmat and the Neuchatel catchments appeared to get slightly larger precipitation events with additional 20 to 40 mm compared to the other sub-catchments. This is even more visible from the return level maps associated to the maximum return periods for the 89 sub-catchments. Similar to the results from GWEX, the higher precipitation values are located on the extreme South-East of the Aare catchment and on the Western part of the area, close to the Jura mountain range. Noticeable are the large differences from one catchment to the other with return levels ranging from 150 to 350 mm for the MAP1d and 200 to 450 mm for the MAP3d. This uneven spatial structure is also visible in the observations for the 150 year return period.

All in all, SCAMP provides relevant meteorological time series that efficiently reflect the climatology of both precipitation and temperature.

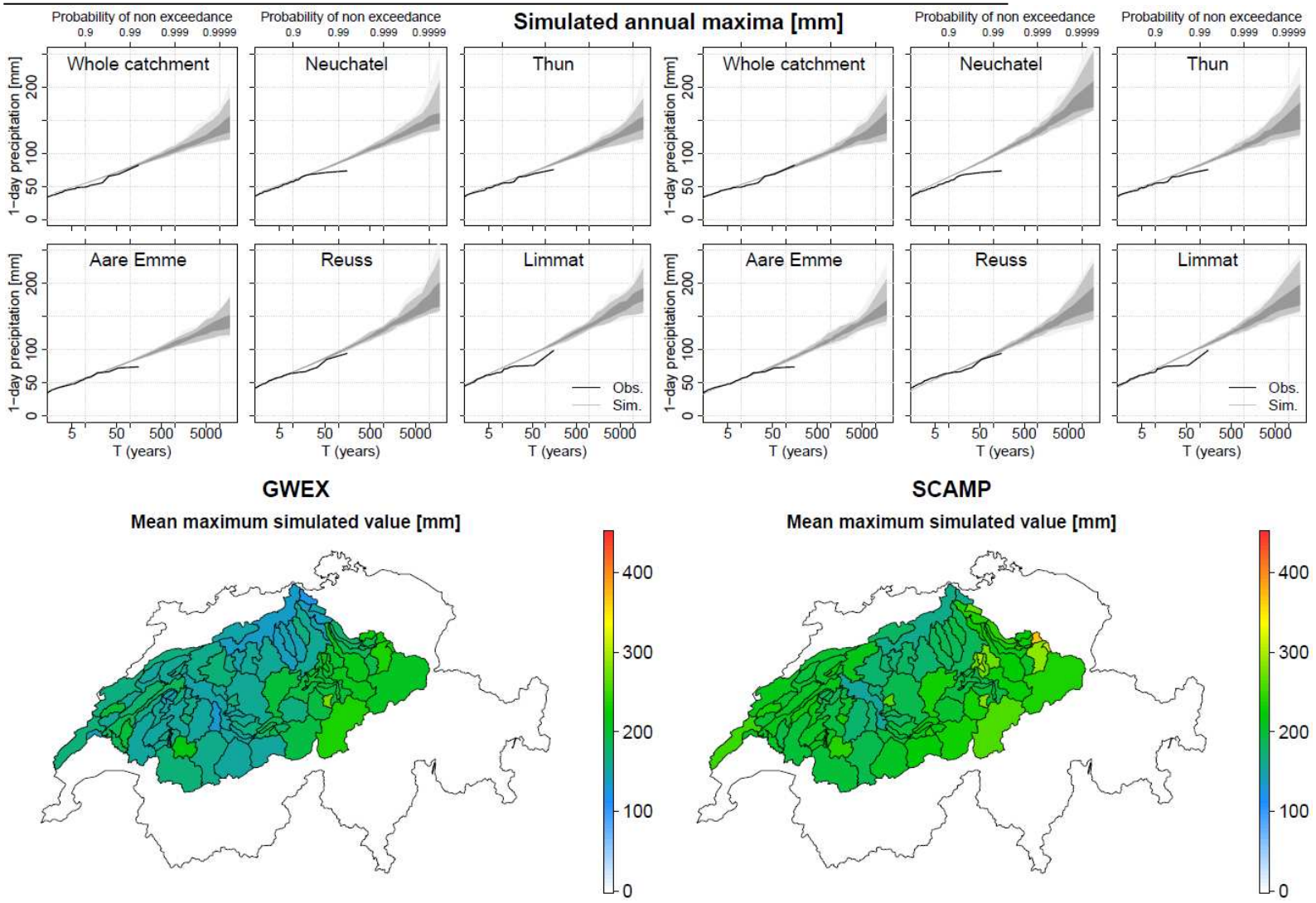
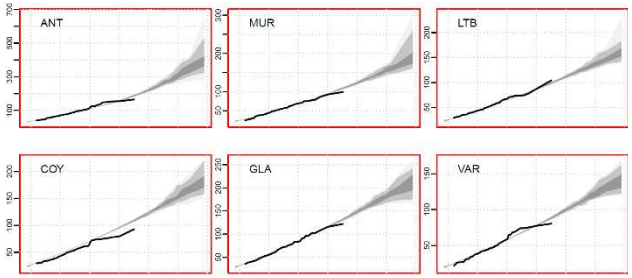
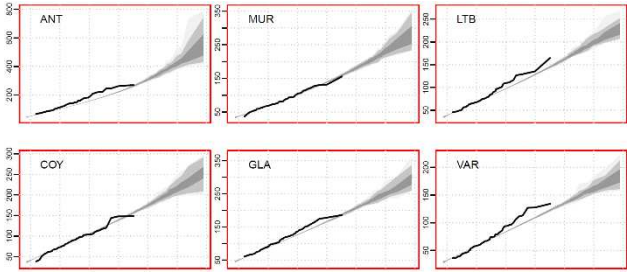


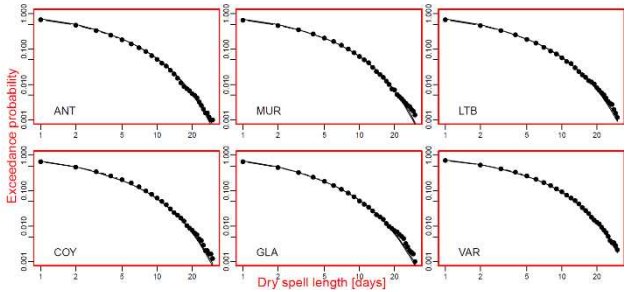
Figure 3 Empirical return levels obtained for MAP1d (1day Mean Areal Precipitation) from the 30 generated time series of 10'000 years for GWEX (left) and SCAMP (right) : Graphs (upper panel) : return levels for the whole catchment and the five main sub-catchments for several return periods (for each return period, mean (middle of the dark grey band) and percentiles of the distribution obtained from the 30 time series scenarios; bounds of shaded areas correspond respectively to 0.5/99.5th, 5/95th and 25/75th percentiles of the 30 time series); Maps (bottom panel): mean maximum simulated value from the 30 x 10'000 year time series for the 89 sub-catchments. Note that the largest simulated MAP value in one 10'000-year-long simulation correspond to a return period of 18'000 years according to the Gringorten empirical plotting position formula. Note that extreme MAP value do not necessarily occur at the same time. Maps of MAP do thus no correspond to ONE extreme event. If locally, for a given subbasin, SCAMP extremes are larger than GWEX extremes, note that MAP value for the whole Aare Catchment are almost the same for the 2 weather generators (see Figure 3 and 5).



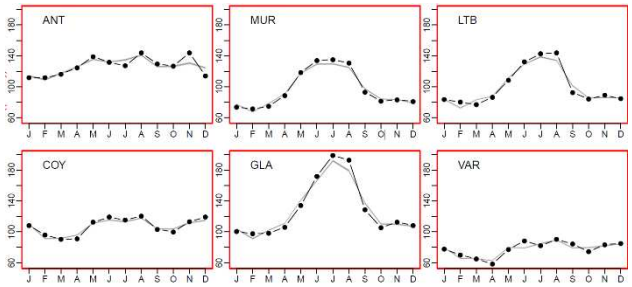
Return levels (1-day precipitation)



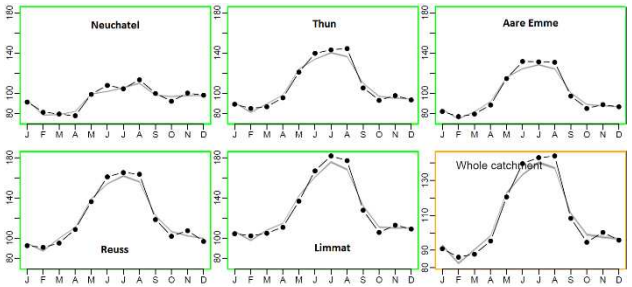
Return levels (3-day precipitation)



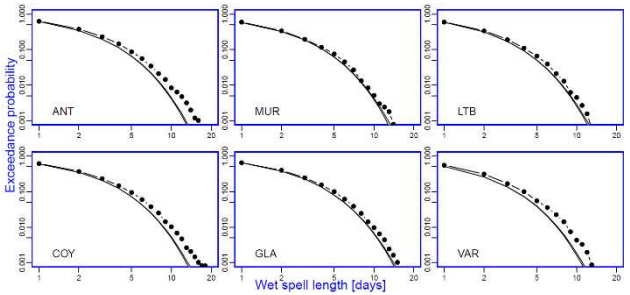
Dry spell lengths



Monthly mean precipitation (6 selected stations)



Monthly mean precipitation (5 large sub-catchments and whole catchment)



Wet spell lengths

Figure 4 Multiscale evaluation of return levels for 1-day and 3-day precipitation, dry spell lengths, monthly mean precipitation and wet spell lengths. The 6 selected representative stations are Andermatt (ANT), Muri (MUR), Lauterbrunnen (LTB), Courtelary (COY), Glarus (GLA) and Valeyres-sous-Rances (VAR).

2.6 Large Precipitation Events in GWEX and SCAMP

At the scale of the whole Aare catchment, MAP extremes are roughly the same for the two weather generators (Figure 3 for MAP1d and Figure 5 for MAP3d). At the sub-catchment scale, SCAMP extremes are slightly larger than GWEX ones for the western part of the system. The largest difference, found for the Neuchatel subbasin, is however moderate (+10% for MAP3d and +20% for MAP1d). An important result is thus that the two weather generators produced similar large/extreme precipitation amounts at all temporal/spatial aggregation scales, despite the very different modelling approaches.

Precipitation time series generated with both models were further evaluated on their severity maps and on their spatial/temporal dynamic and their meteorological relevance. The GWEX precipitation event that led to the largest simulated flood of the Aare is illustrated in Figure 7 and Figure 8. Similar figures are presented for other large events in Appendices 10.5.2 and 10.5.3 for GWEX and Appendices 10.5.4 and 0 for SCAMP. The severity map of the event for the full Aare catchment, i.e. the severity of the event in terms of return period of cumulated precipitation amounts for different locations, spatial and temporal scales as well as its spatial/temporal dynamics over the 10 days preceding the flood peak are shown in Figure 7.

For both GWEX and SCAMP, a large variety of severity maps and of spatial/temporal dynamics were generated in accordance with the large variety of events present in the observation period. The largest floods from the generated time series came from very different hydro-meteorological configurations. Some of them corresponded to very large precipitation amount at the whole catchment scale for one or two days previous to the flood peak; some of them conversely corresponded to very large precipitation amounts falling in a small part of the whole catchment or to configurations with large precipitation amounts that fell during a few days preceding the flood. These different configurations have been estimated to be plausible by MeteoSwiss (personal communication with Sophie Fukutome and Andreas Fischer).

Note that the largest generated values obtained with GWEX and SCAMP were compared to precipitation values estimated in a former project with the Probable maximum Precipitation (PMP) method (Hertig and Fallot, 2009; Appendix 10.6). GWEX and SCAMP maximum precipitation values are smaller than PMP values. Note that this comparison was carried out for information only. PMP estimates are local and cannot be provided for medium size to large size catchments (i.e. for scales relevant for the Aare catchments and for its main sub-catchments). The PMP method is also based on a meteorological model which is uncertain itself and based on a set of different hypotheses and sub-grid parametrizations. PMP estimates were also made for a small sub-set of possible large scale atmospheric configurations. This makes the PMP estimates obtained within this project rather uncertain and in all cases dependent of the modelling and parametrization choices made for the application of the method. As no uncertainty analysis was provided within this PMP work, any quantitative comparison of PMP estimates with maximum precipitation amounts obtained from GWEX and SCAMP scenarios is therefore neither possible nor relevant.

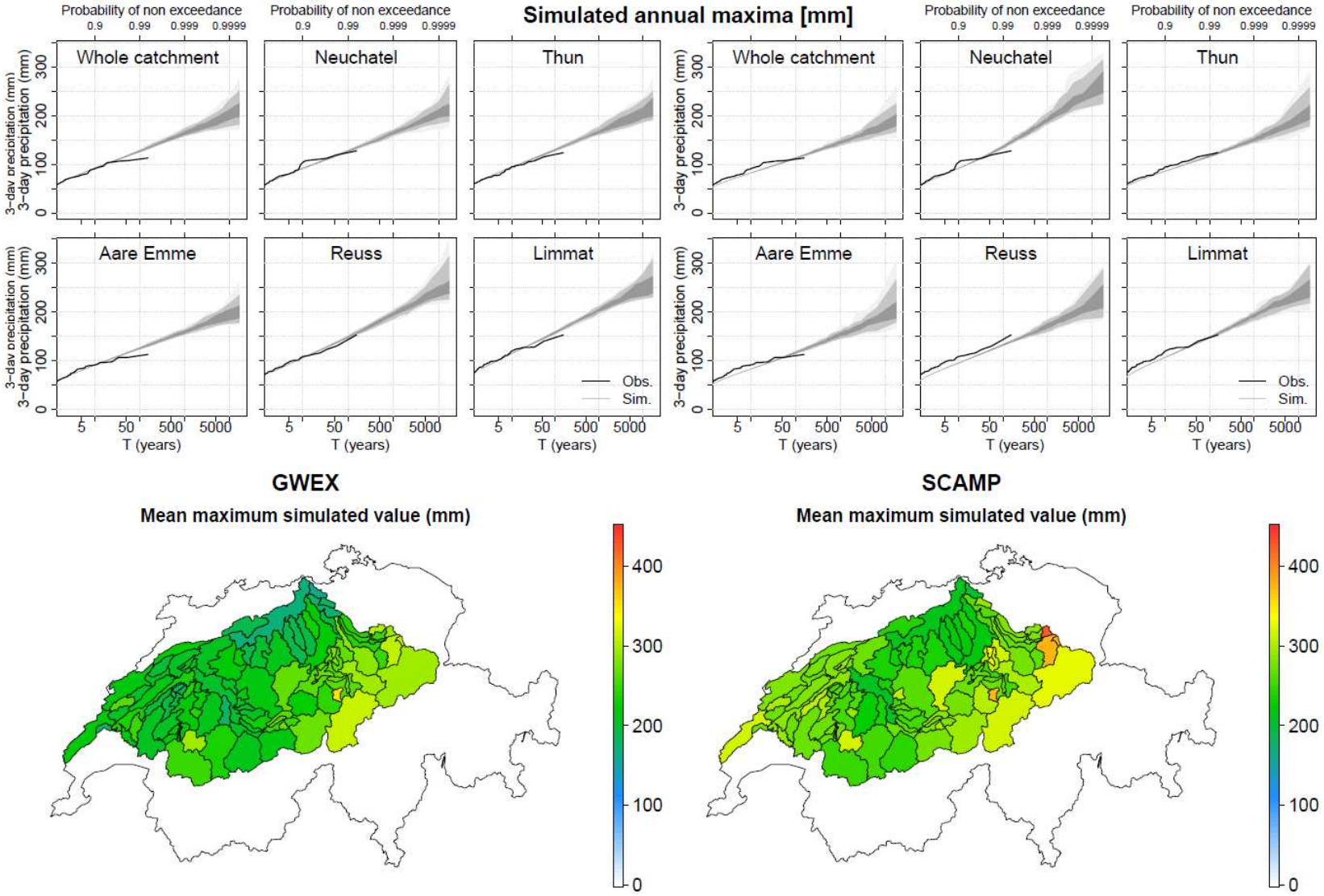


Figure 5 Same as Figure 3. Precipitation maxima for MAP3d.

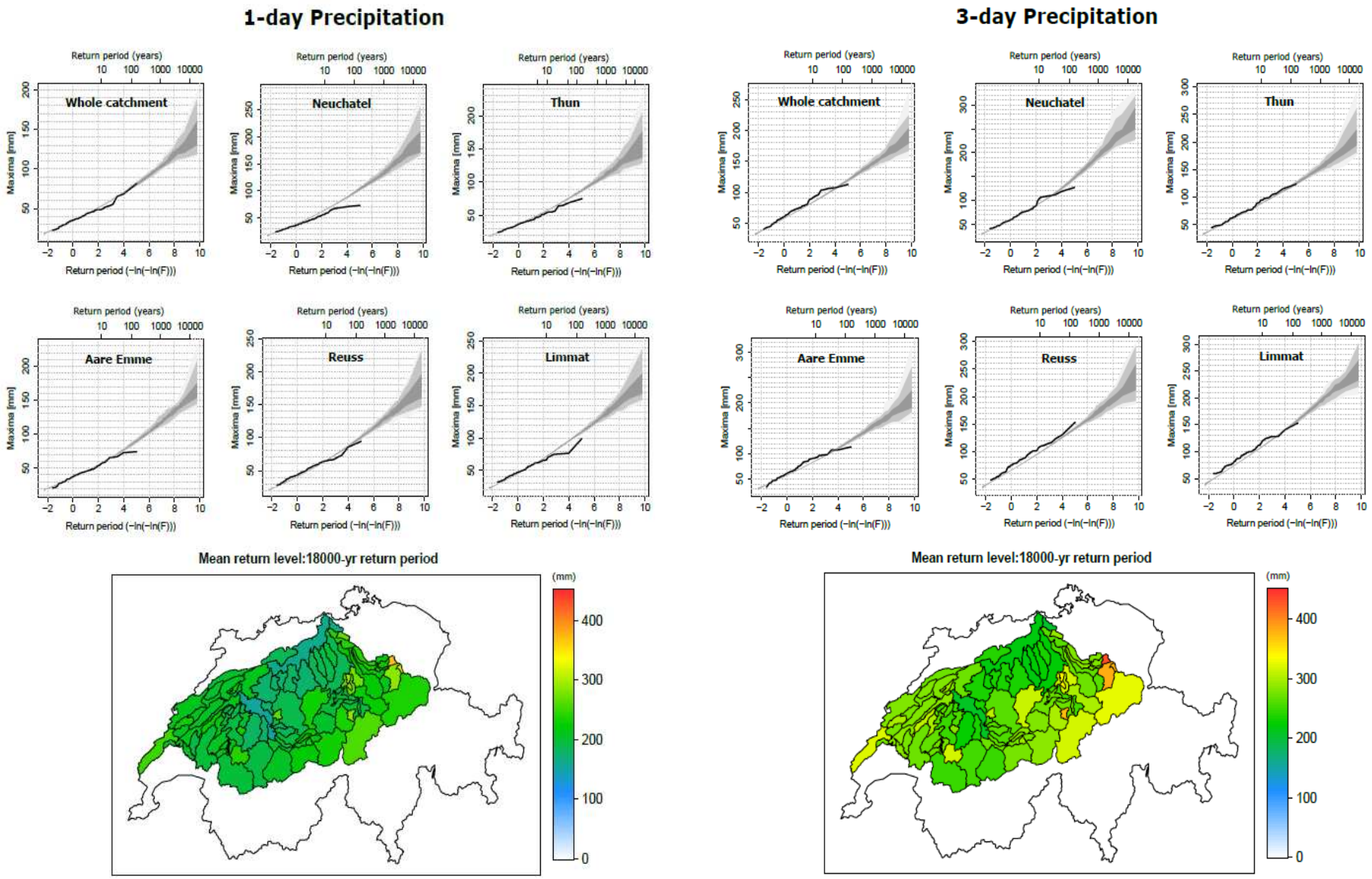


Figure 6 Same as Figure 4. Precipitation maxima for SCAMP weather scenarios

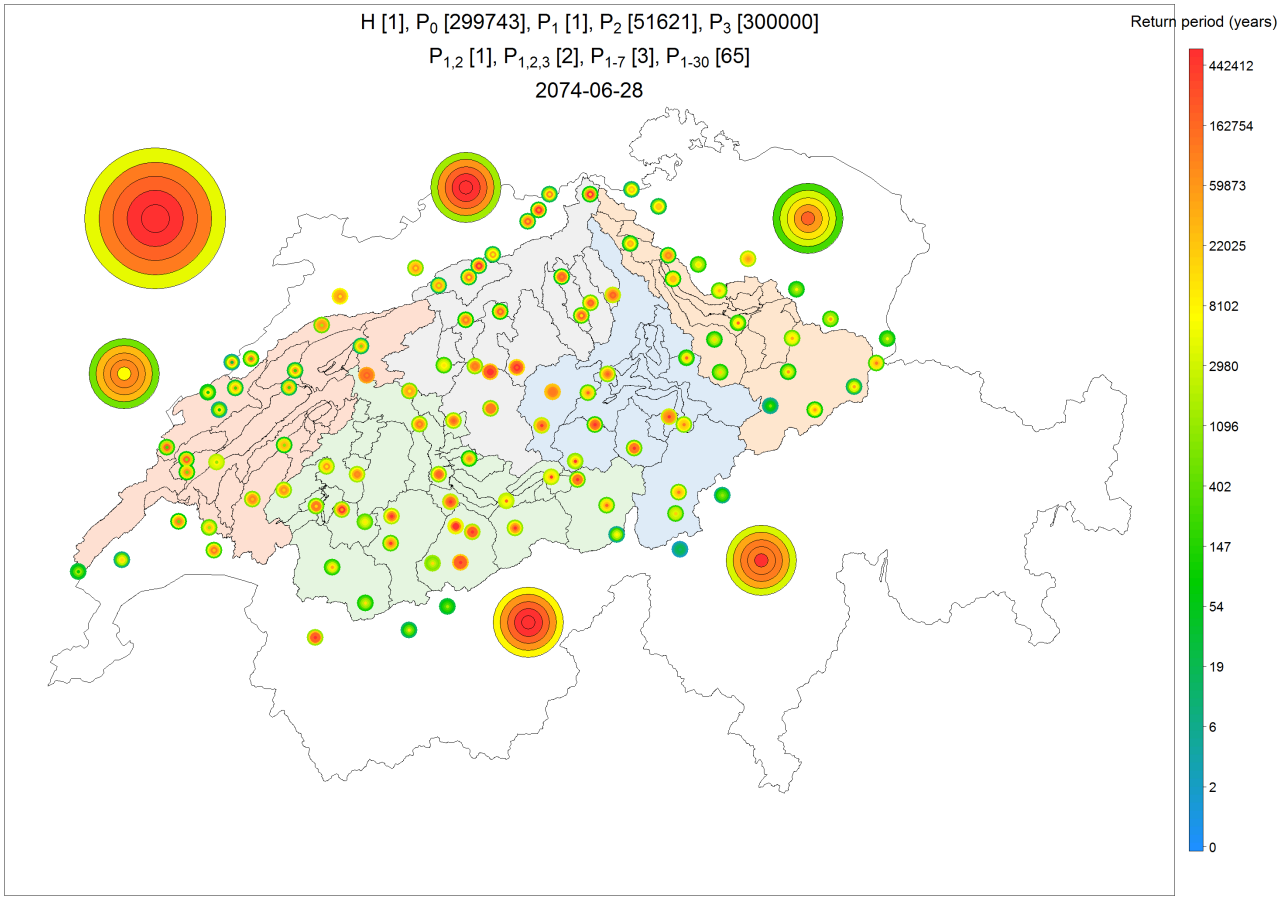


Figure 7 Severity map of precipitation during the largest hydrological event. The title shows the rank (in descending order) of the precipitation events at the whole catchment scale for several accumulation periods (e.g., P0[299743]: the precipitation during the day of the hydrological peak is the 299743 th1-day precipitation amount (i.e., a very frequent precipitation amount); P1[1]: the precipitation during the day before the hydrological peak is the first 1-day precipitation amount (i.e., the largest 1-day amount simulated with GWEX over the 300'000 years of simulation); P1,2[1]: the cumulative precipitation during the two days before the hydrological peak is the largest simulated 2-day precipitation amount). The severity of the simulated event is then represented with a set of severity-board. Each severity-board gives the return period of precipitation simulated for this event when cumulated over several durations (all ending the day before the hydrological peak): 1 day (center of the severity-board), 2 days, 3 days, 7 days and 30 days (edge of the severity-board). The color scale for the severity-boards is given in the right hand side of the figure. Severity-boards are given for precipitation simulated over different spatial scales: local precipitation at each of the 105 simulation stations (small circles), mean areal precipitation over each of the 5 main sub-catchments (medium circles) and the whole catchment (large circle).

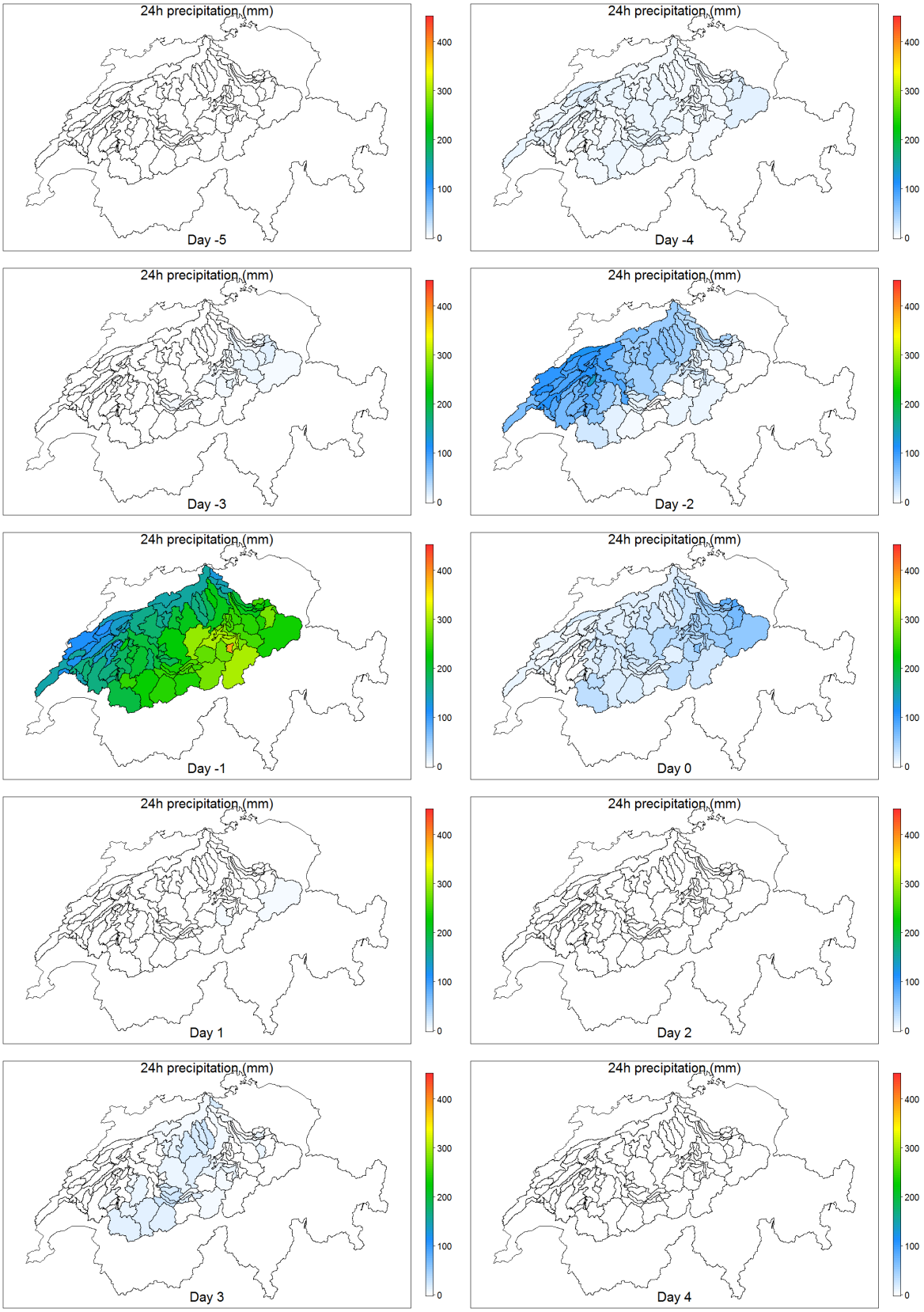


Figure 8 Space-time dynamic of 1-day precipitation during the largest hydrological event. "Day 0" is the day of the hydrological peak.

Note: Absence of plateau in the tail of the distribution of peak discharges. As a result of the EXAR methodology, there is no reason why we should have a plateau. Except in some specific configurations (which we do not have in EXAR), it is expected that the peak values keep getting larger for small probabilities of exceedance, and it is also expected to have some "super events". We would even have larger super events if we would generate scenarios over longer time periods (>>300'000yrs). This directly derives 1) from the basic principles of the simulation chain we choose to develop for the project (weather generator + hydrological model) and 2) from the statistical behavior of precipitation in the Aare region.

As already mentioned, GWEX weather scenarios are generated to reproduce the statistical behavior of precipitation data observed over 105 Aare stations for the last 85 years. The behavior of observed precipitation maxima in our Aare basin does not show any plateau or any asymptotic trend for a plateau. Therefore, there is no plateau in precipitation scenarios. Even if the rainfall-runoff transformation is not linear, it is also not surprising that there is also no plateau in hydrological scenarios.

The uncertainty on the precipitation maxima values for the lowest probabilities is very high (due to extrapolation for a 300'000 year return period from distributions obtained with observations covering only 85 years). A critical issue here is that of the estimation of the shape parameter of the distribution used to model the statistical behavior of observations. The value obtained for this parameter defines the behavior of the largest precipitation discharge generated values. As we used advanced regionalisation techniques for this estimation, we are rather confident in the robustness of those estimates.

It is important to put the outputs of GWEX simulations in perspective with those obtained with SCAMP, a fully alternative modelling approach we chose for weather scenario generation. Concerning the plateau, the same applies for weather scenarios generated with SCAMP as for GWEX. Even if the basic principles are different from those of GWEX, SCAMP weather scenarios are generated to reproduce the statistical behavior of observed precipitations. And, as no plateau is found in the observed data, no plateau is expected in generations.

3 Sensitivity and plausibility of hydrological simulations

Martina Kauzlaric, Luise Keller, Matthias Pfäffli, Anna Sikorska-Senoner, Maria Staudinger, Daniel Viviroli

This chapter provides background to the hydrological simulations, covering both the HBV model and RS Minerve model. Focus of the analyses is the sensitivity of the hydrological models to the input from the weather generators as well as to the parameters used for regionalization, model performance with regard to discharge observations as well as the plausibility of the simulations in terms of spatial patterns, seasonality and temporal development. The IDs for the analyzed catchments as well as the transfer points are listed in Appendix 10.1.

3.1 Sensitivity of the hydrological models (HBV and RS Minerve)

The sensitivity of the catchments simulated with HBV was assessed by two tests regarding two different aspects:

1) **Uniform precipitation input** was used as model input, and the resulting specific runoff peaks were evaluated. This analysis served to assess the realism of the runoff yield HBV model and its sensitivity towards precipitation inputs. This analysis allows an insight into the impacts of meteorological versus hydrological simulations on the generated extreme runoff peak values.
2) **Different spatial discretization in the floodplains** was used to assess the influence of the spatial discretization of large ungauged sub-catchments located in the floodplains of the major rivers (Aare, Reuss, Emme and Limmat) on the peaks. The corresponding catchment areas were divided into smaller units, and in separate parameter sets for each of these units were used.

For the first sensitivity test, the results of each individual catchment as well as at important transfer points (TPs) were compared; for the second test the results were compared only for the TPs, as the higher spatial discretization concerned only the floodplains and not the headwater catchments.

3.1.1 Simulations with uniform precipitation

To assess the sensitivity of the simulations to the hydrological model rather than the differences in meteorological input, uniform precipitation was used as model input, i.e. for each catchment the same amount of precipitation was applied at the same time. The uniform precipitation was derived from the sum of precipitation inputs from individual catchments that were then averaged over the entire Aare system. The temperature input remained unchanged. These uniform simulations with uniform precipitation were computed for two variants:

- Using 85 years (1930–2014) of pseudo-observations, i.e. disaggregated precipitation values
- Using the 10 largest events extracted from 289’000 years of simulations based on GWEX weather generator input.

Both simulations were done using the parametrization resulting from all three clusters of the HBV model calibration. However, here the results are shown and discussed for one cluster only.
For the pseudo-observations, annual runoff maxima were extracted from uniform simulations in each catchment, and the three events with the largest precipitation peaks, which were most frequently classified across all sub-catchments, were further analysed in detail. The three largest precipitation events in the pseudo-observations occurred in 1954 (simulation year 25), 1978 (simulation year 49) and 1990 (simulation year 61). Further, simulations were done for the six largest events in terms of discharge peaks and another six based on the largest events in terms of discharge volumes. For these events all catchments were analysed.

For the GWEX simulations, the 10 largest events were analysed in detail. For this purpose input data covering the time span of 1.5 years before the flood event and an additional 10 years period for model warm-up were extracted from the input time series and averaged over the entire Aare system. This average formed the uniform precipitation inputs for the GWEX simulations. These data were then used as input for the hydrological model and simulations were done for the three parameterizations (lower, median upper): For the 10 largest events HBV simulations were run using a 10 years warming-up period prior to the year when the largest event occurred and 1.5 year of simulations for the largest event.

For both the pseudo-observations and the GWEX simulations metrics were computed for the selected events (Table 1). For the pseudo-observations, the metrics were computed for each catchment independently, i.e., in accordance to the time of the annual peak occurrence. Despite the uniform

precipitation input, the runoff peaks may still occur at different times because of the reaction time of the catchment caused by different storage behaviour in each catchment. With these metrics the events can be characterized and the catchments can be compared in a spatial context.

Table 1 Metrics to analyse the pseudo-observations and GWEX simulations for the uniform precipitation simulation experiment.

Variable	Metric	Unit
Discharge	Runoff peak of the event	(mm/h)
Precipitation	Maximum amount of precipitation preceding the event peak	(mm/h)
Discharge	Maximum annual sums of runoff over a moving window of 72h and 144 h	(mm)
Precipitation	1, 3, 5 and 10 days sums of preceding precipitation to the 72 and 144 runoff peak events	(mm)
Discharge and Precipitation	Runoff ratios (runoff/precipitation) of the 72 and 144 h runoff sums to 1, 3, 5, and 10 days sums of preceding precipitation	(-)

Pseudo-observations largest events

Figure 9 shows the preceding precipitation for an aggregation window of 120 h for the three analysed years as selected from the largest precipitation events. While in 1954 and 1978 mainly different storage characteristics led to different antecedent precipitation periods taken into account for the analysis, in 1990 there is a clear difference between the high alpine catchments and the rest of the Aare system, which points at a strong influence of snow melt on the highest peak runoff in this year.

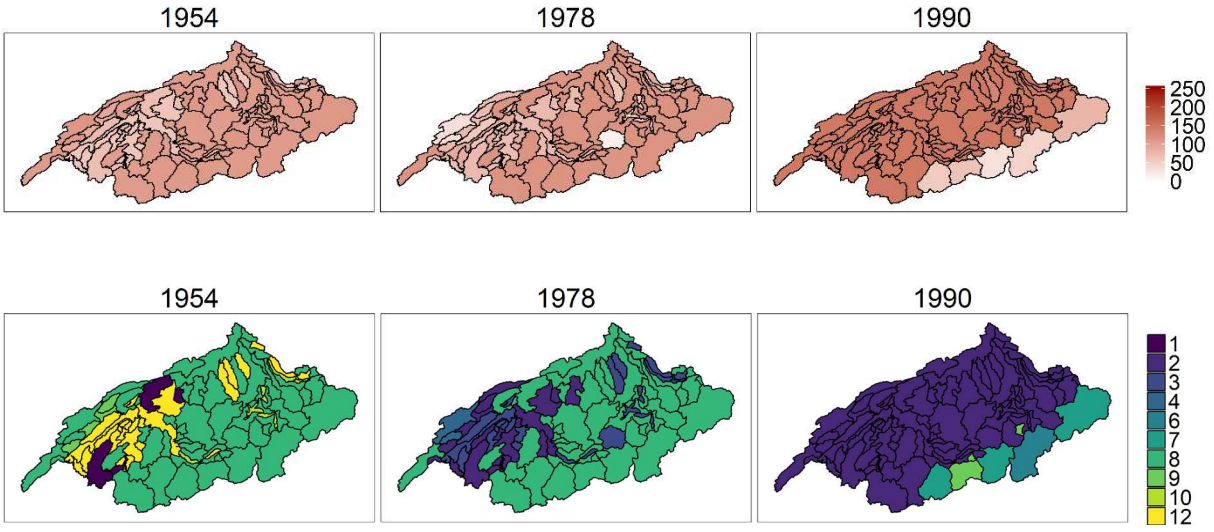


Figure 9 Precipitation sums for a window of 120 h preceding the largest runoff peak in the respective year (upper panel) and the month of occurrence of the peak runoff per catchment (lower panel). All values are given in [mm].

The runoff ratios of both 72 h and 144 h time windows are larger than 1 for some catchments (Figure 10), which indicates that there was more runoff than preceding precipitation. This can occur when other sources than precipitation contribute to runoff, for instance snow melt as a delayed contribution. In 1990 the peaks occur in the spring and early summer months and hence snow melt can indeed

explain the high runoff ratios. For other years the reason for runoff ratios above 1 might be that the catchments concerned are located in karstic areas, where the model might not be adequately parameterized (see also Section 3.6.2 about the performance of hydrological models for each catchment). Most of the karstic catchments were parameterized by regionalization, which had to make use of donor catchments that are not karstic. The donor catchments do not account for the loss of water that we would find in karstic systems, which can result in unrealistically high runoff ratios.

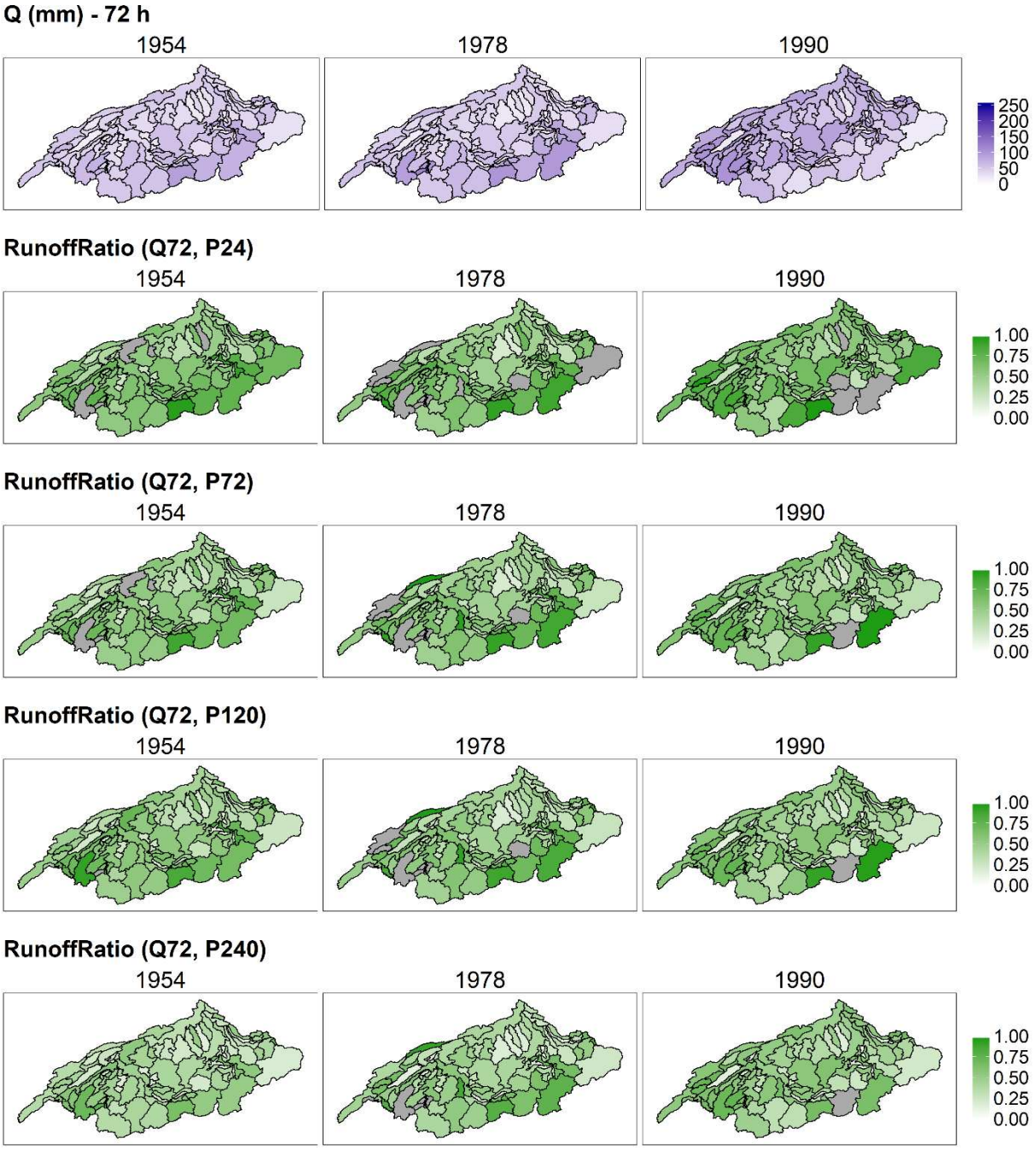


Figure 10 72 hours runoff sums (uppermost panel) and the runoff ratios resulting from the preceding precipitation and this runoff for different aggregation periods for the precipitation. The grey colored catchments have a runoff ratio larger than 1, meaning that there was more runoff than precipitation. All values are given in mm.

The longer the aggregation window for the antecedent precipitation sums were chosen the fewer catchments had runoff ratios above 1. The spatial distribution of runoff ratios is not changing much for increased aggregation windows of precipitation, which indicates that the events of precipitation were really intense and happened mainly right before the highest peak of the runoff.

Very similar spatial patterns were also found for the event aggregation of 144 h for runoff. For the longer aggregation period, the runoff ratios are above 1 also for longer precipitation aggregations for some catchments, simply because the runoff sums are naturally larger over the longer aggregation time of the runoff time series for 144 h compared to 72 h.

For some events, rain-on-snow processes might contribute to large floods. Such a rain-on-snow event could have been induced in the simulations with uniform precipitation (Figure 9) for the year 1990. Here, the Alpine catchments behave differently to the rest of the catchments, which indicates the influence of snow, maybe also rain-on-snow. The hydrological model has not been designed to reproduce the physical processes occurring during rain-on-snow events, and snow melt floods that occurred during such an event might be larger than simulated without explicitly accounting for the processes during rain-on-snow. The duration of the considered extreme events, however, is long enough (72 hours) to allow for a high snowmelt contribution even in absence of a special “rain-on-snow” parametrization. Furthermore, the largest events that we found in the GWEX simulations (not in the sensitivity analysis using uniform precipitation) occurred in seasons where we do not expect rain-on-snow events (see Figure 65), and they have very large 72h precipitation sums (Figure 64) that reach 400 mm and more for some catchments.

From this analysis can be concluded that the largest precipitation events in the sub-catchments demonstrated consistent spatial characteristics, although slight differences were observed between individual sub-catchments. Although all sub-catchments received exactly the same amount of precipitation at the same time, runoff peaks did not occur at the same time. Thus, time responses were different in these sub-catchments. Also, runoff peaks and sums were different. Sub-catchments located close to each other or with the same properties generated similar runoff amounts during the flood event. Such spatial coherence was observed in runoff sums and ratios at larger aggregation windows, while short aggregation windows lead to a more heterogeneous pattern of the responses of the sub-catchments. The effect of the temperature that was kept unchanged for each catchment for the sensitivity analysis is still visible since there are some peaks that were created from snow melt and not precipitation alone. The snow melt contribution to peak flow resulted in runoff peaks that did not occur at the same time of the year for the entire system but forced a seasonal separation in the spatial patterns.

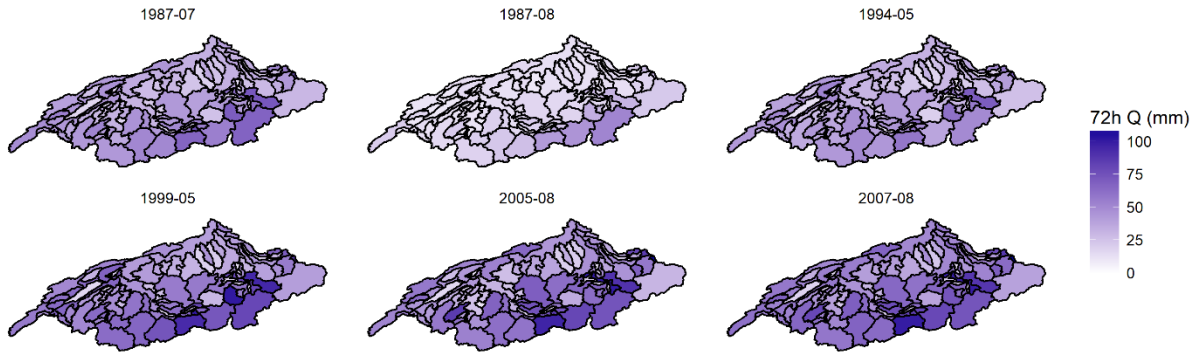


Figure 11 Pseudo-observations: Discharge sum over 72 hours for the largest peak events for uniform precipitation inputs.

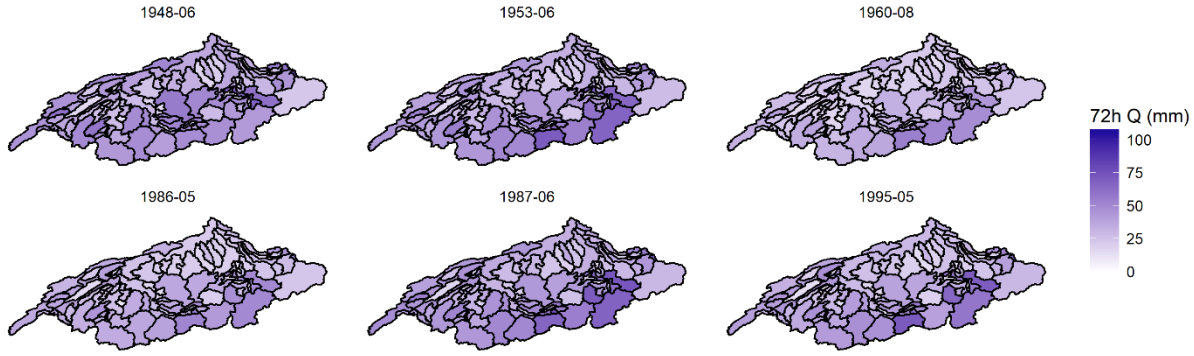


Figure 12 Pseudo-observations: Discharge sum over 72 hours for the largest volume events for uniform precipitation inputs.

The largest events in the pseudo-observations with respect to the entire Aare system were selected in terms of peak flow (six events) and discharge volume (six events). For these largest events we calculated the runoff peaks and sums (72 and 144 h) and antecedent precipitation sums (1, 3, 5, 10 and 28 days).

The largest volume and largest peak events from the pseudo-observations reached very similar sums over aggregation periods of 72h for some catchments (Figure 11 and Figure 12). For the largest peak events however, the largest sums are found distinctly in the alpine catchments, whereas for the largest volume events this is more distributed over the entire system. Neighbouring and similar catchments have a similar reaction on the precipitation input. The same patterns were also found for an aggregation period of 144h.

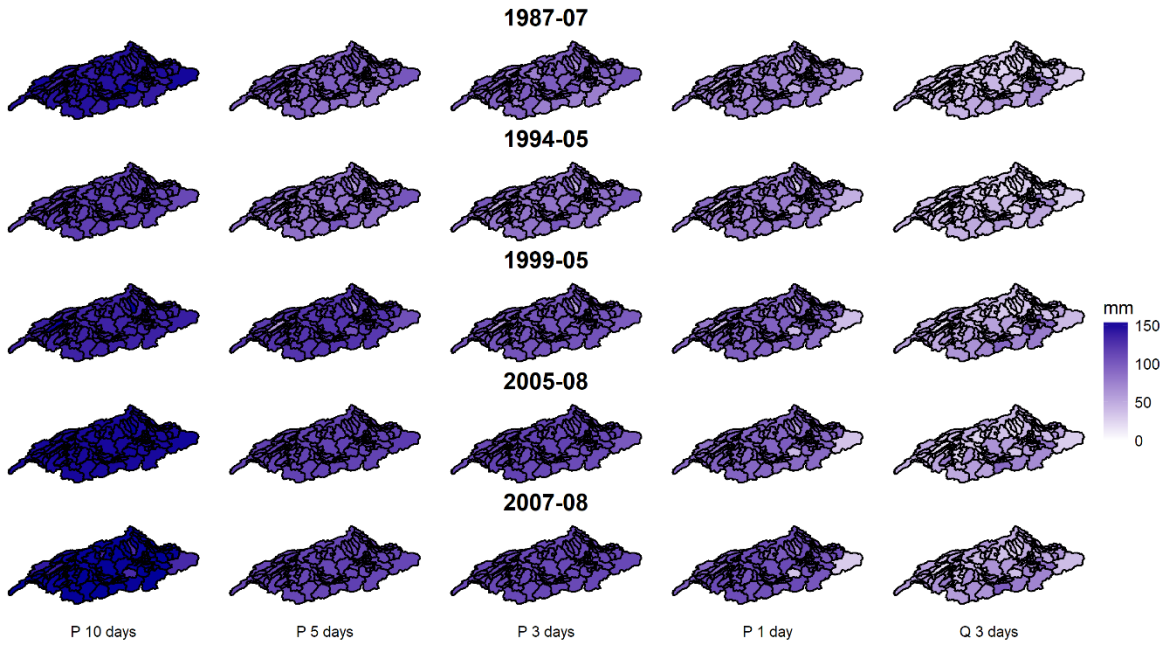


Figure 13 Pseudo-observations: Antecedent precipitation at different aggregation windows and 3 days (72h) runoff sum for largest peak runoff events.

The runoff response of the catchments is relatively diverse compared to the very similar precipitation input (Figure 13). The largest differences are visible in the one day aggregations, and, less pronounced, in the three day aggregations. The difference, which is basically due to the timing of the peak runoff, is then only visible for a few catchments for longer aggregations. The runoff reactions of neighbouring catchments are similar and most of the diversity observed is found in the alpine catchments.

GWEX simulations largest events

The uniform precipitation inputs derived from the GWEX simulations for the largest ten events are summarized in Table 2. For most of these largest events the major part of precipitation is already falling on the first day before the event. This is less pronounced for the precipitation derived from data series 58, 111 and 253. There, the distribution of the precipitation event is more spread over the days.

Table 2 Characteristics of the precipitation sums for the largest events, sorted from the largest simulated data series (#96) to the smallest (#253).

Data series	Precipitation uniform input			
#	1 day (mm)	3 days (mm)	5 days (mm)	28 days (mm)
96	161.8	177.5	188.0	300.4
271	69.4	77.0	114.3	185.7
138	138.1	142.9	146.8	329.0
286	164.6	185.3	195.6	280.1
38	137.2	160.6	190.4	257.7
291	48.5	73.1	86.0	189.3
58	127.5	175.6	191.6	314.0
111	60.0	63.0	85.3	135.6
158	58.3	58.3	63.9	212.7
253	84.3	114.3	128.0	210.8

The simulations for 3 largest events (data series 96, 271 and 138) and the median cluster are presented in more detail (Figure 14). To assess the effect of antecedent precipitations sums, these largest events are set in relation to their antecedent precipitation and with increasing aggregation windows. A spatial smoothing can be found, i.e. antecedent precipitation patterns become more uniform the longer the aggregation time is.

Despite the uniform precipitation input that was applied over the catchment models the variability in the runoff response is large. For the two largest events (data series 96 and 271) the variability of the runoff response is even higher than for the third largest event (data series 138), which is because many of the alpine catchments react stronger to the large uniform input. With less overall precipitation input also the response became more uniform.

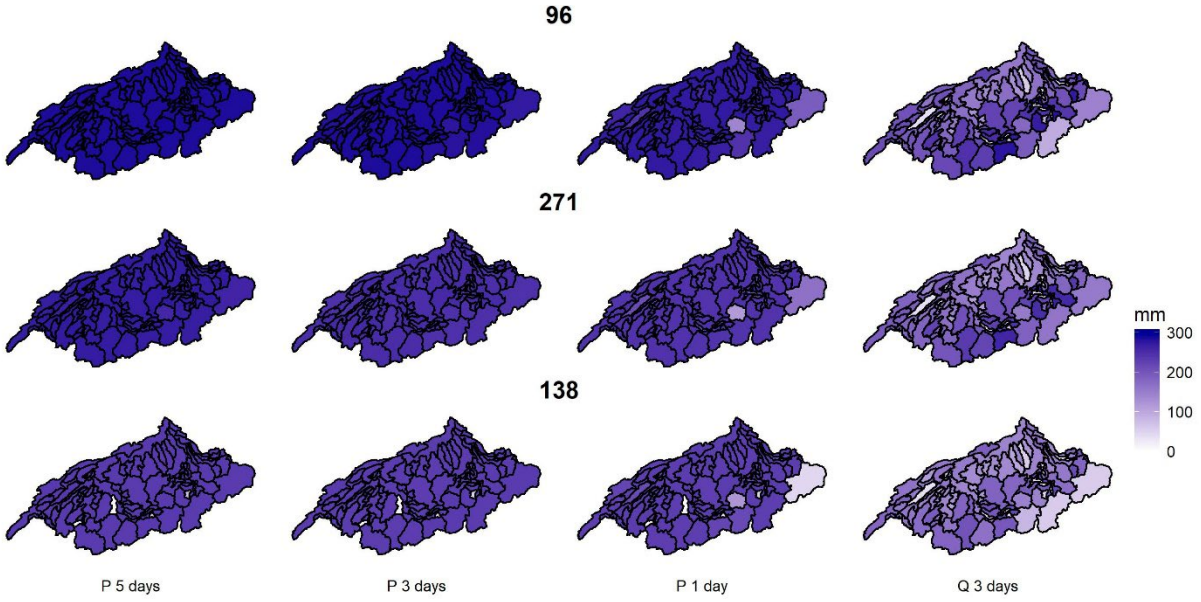


Figure 14 AMFs for a 3 days window compared to the amounts of antecedent precipitation for different aggregation periods.

The discharge obtained by HBV with spatially uniform precipitation time series for the 10 largest GWEX events was further fed to RS Minerve, in order to see the overall effect on the TPs. No general pattern across all TPs was found: e.g. while at the outlet of lake Lucerne there is a decrease of 15 to 40% of peak discharge for most events with 2 exceptions, at the outlet of the lake of Thun all but three events show an increase in peak discharge, however the differences are smaller than those observed in the first case. The results obtained for the outlet of the catchment are shown in Figure 15, where positive deviations are usually slightly higher than negative deviations from the original results. It is worth to note that differences among different parametrizations are often higher than the differences obtained by applying different precipitation time series.

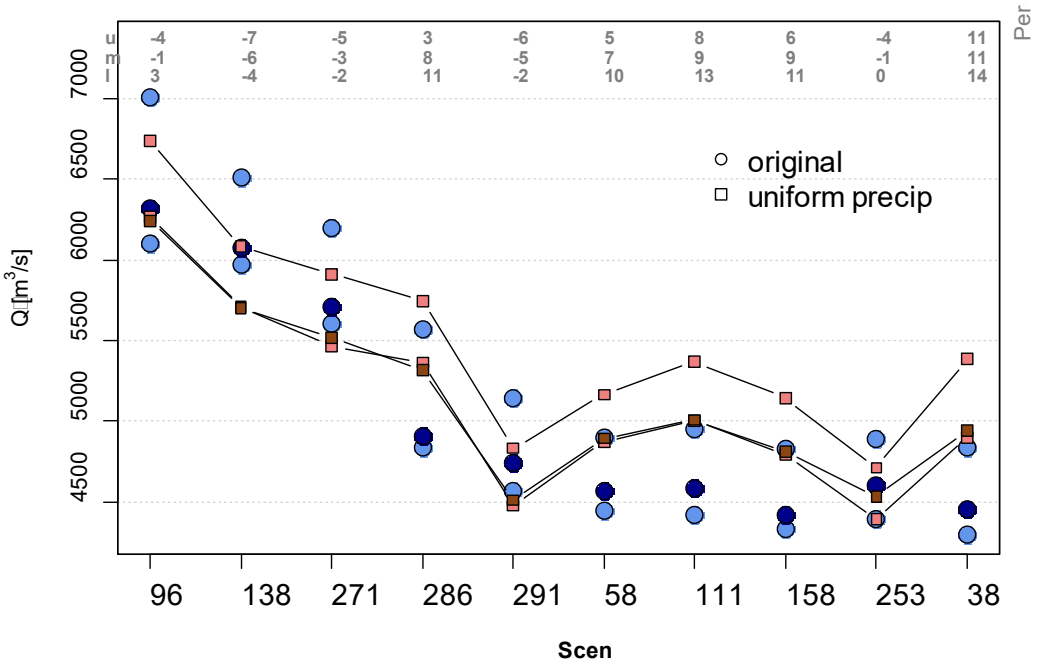


Figure 15 Peak discharge of the 10 largest events at the outlet of the catchment (SSIOut) applying a uniform precipitation distribution (squares) compared to the original GWEX results (circles). The median parameter set highlighted by a darker colour. At the top of graph the percentage change for each event and parameter set is indicated.

3.1.2 Spatial discretization in the floodplains

In the original GWEX simulations a pragmatic approach was taken to model the ungauged floodplains of the major rivers (Aare, Emme, Reuss and Limmat) that consisted of modelling them as one big entity in the HBV model, and feeding it to the system as a whole at the outlet of the floodplain. E.g. the floodplain of the Aare between Thun and Hagneck (inlet to Lake Biel) was fed into the total system directly in Lake Biel. To test the effect of a different approach, we divided these large floodplain sub-catchments into smaller sub-units. The total floodplain discharge was first area-weighted to obtain discharge for these sub-units, then it was diverted and fed in the total system at appropriate locations along the main river.

We performed simulations applying this approach for the ten largest events. Accordingly, larger differences are simulated for TPs which did not receive floodplain discharge before (e.g. SSASSB, SSBSSC), showing a systematic increase of peak discharge. The magnitude of this effect depends on the event, on the discharge contribution of other tributaries along the main river as well as on the presence of retention areas in the floodplain. The overall effect at the outlet of the catchment is rather small, with a tendency towards smaller runoff values (see Figure 16).

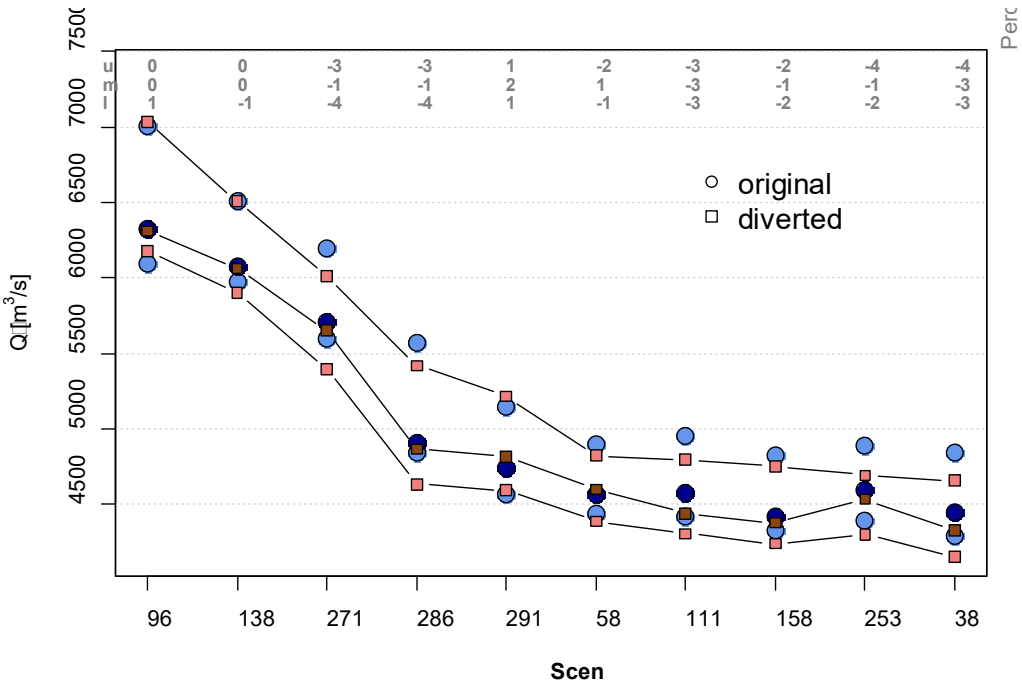


Figure 16 Peak discharge of the 10 largest events at the outlet of the catchment (SSIOut) simulated by diverting discharge of the floodplains (squares) compared to the original GWEX results (circles). The median parameter set highlighted by a darker colour. At the top of graph the percentage change for each event and parameter set is indicated.

By regionalizing and implicitly re-parametrizing the floodplains' sub-units the differences compared to the second more simple approach diverting discharge are negligible (see Figure 17).

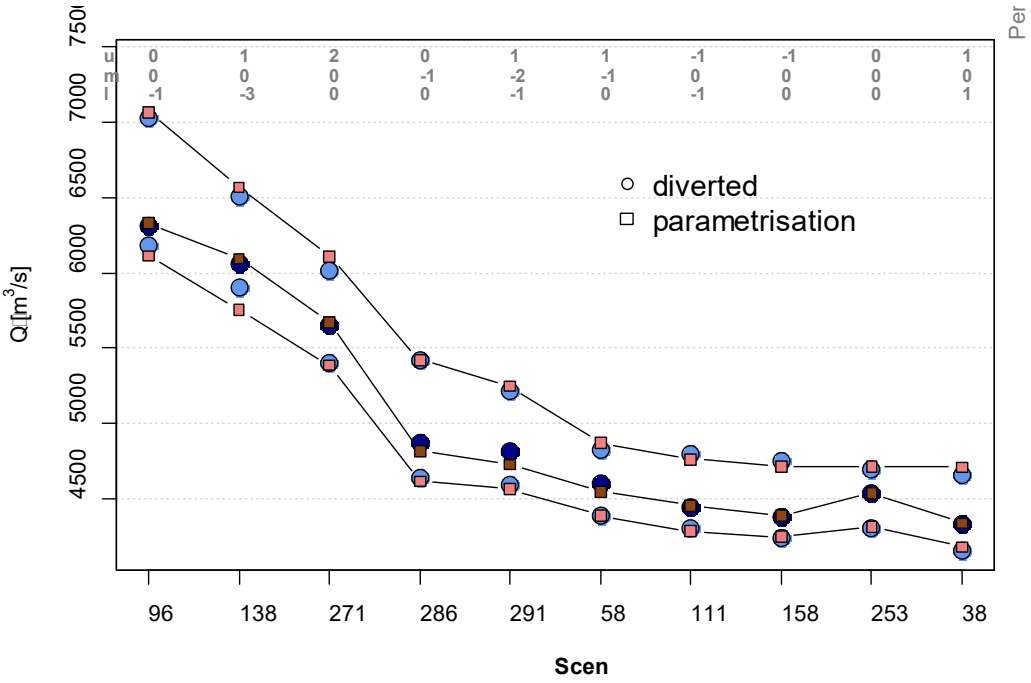


Figure 17 Peak discharge of the 10 largest events at the outlet of the catchment (SSIOut) simulated by re-parametrizing the subunits of the floodplains (squares) compared to the original GWEX results (circles). The median parameter set is highlighted by a darker color. At the top of graph the percentage change for each event and parameter set is indicated.

3.1.3 Conclusion on sensitivity of hydrological models

Concerning the pseudo-observations, the runoff response of the catchments is relatively diverse compared to the very similar precipitation input (Figure 13). Regarding aggregation time, the largest spatial differences are visible in the aggregation over one day, and less pronounced in the aggregation over three days. For longer aggregation times, the difference, which is basically due to the exact timing of the peak runoff, is visible only for very few catchments. The runoff reaction for neighbouring catchments is similar and the main largest diversity is present in the alpine catchments.

Concerning the GWEX simulations, despite the same inputs (precipitation) the sub-catchments demonstrated different behaviour regarding the magnitude of the runoff events and runoff ratios. Neighbouring sub-catchments or sub-catchments with similar properties (catchments with glacier cover, catchments in the Jura region, and lake tributaries) demonstrated similar behaviour. Aggregating runoff sums over longer periods led to more uniform runoff sums and runoff ratios in the Aare River basin, while aggregations over a short period (1 day or peak) resulted in more heterogeneous patterns.

It can be concluded that even for the largest events, assuming that the precipitation time series has a uniform spatial distribution, heterogeneity in the sub-catchment responses can be observed. This is particularly pronounced for instantaneous values (peaks) or aggregations over a short period (one day). This means that the catchments' storages were never completely filled, so that any uniform precipitation event would have led to the same reaction time and runoff amount. See also the analysis concerning the initial soil moisture status before large events (3.7.5).

3.1.4 Conclusion on different spatial discretization

While the effect of a refinement of the representation of the floodplains can play a role for single TPs, with differences that are in the order of magnitude similar to those simulated with different precipitation time series (heterogeneous versus homogeneous), this is less distinct at the outlet of the catchment. Applying a new parametrization to the sub-units affects only to a very small extent the overall result compared to the diversion-approach. However, it is questionable if this would still be the case if a different regionalization method was applied and if a larger pool of gauged sub-catchments with lower mean altitudes was used.

3.2 Further analyses on regionalization

As noted in the main report, a number of approaches was tested for finding a suitable regionalization scheme. The basis for the approaches tested were

- metrics of hydrological similarity, both defined in physical space (nearest neighbor) as well as in catchment attribute space (Euclidean distance),
- clustering (k means), and
- random forests.

The number of donors varied between 2 and 5 depending on the application of these methods, selecting parameter sets randomly or selecting only the best ones according to the model efficiency criteria, as well as considering further attributes, such as the climatic region or the runoff regime.

We tested all approaches by declaring sub-catchments where the calibration was poor to average (i.e. multi-objective efficiency criterion <0.4 and 0.4–0.5, respectively) as pseudo-ungauged, in order to

evaluate their performance, and treated glacierized basins separately. Of the 84 non-glacierized sub-catchments, 52 are gauged; 28 of these were retained as potential donors, 4 were removed because of very short or questionable time series of observations, and 20 were used as pseudo-ungauged. Unfortunately, no method in particular lead to promising results. It appears that some (few) donors are repeatedly selected as being the most similar for many catchments even if those are far away, what is probably because many of the target catchments are small and thus are most similar to the few small donor catchments. Additionally, most sub-catchments where the calibration was not successful are not pluvial, conversely to the donors, which are mainly nivo-pluvial to nival sub-catchments. We ran simulations for all pseudo-ungauged catchments, selecting 20 parameter sets (evenly distributed between the best and the worst parameter set) from all 28 donors, and found that for some catchments only 1 donor could produce results comparable to the calibration, that generally 2 donors are sufficient and even the best option, and that for some of the sub-catchments the best donors were simply never selected by any of the methods tested. From these results we derived what would have been the "ideal" clustering, i.e. with 2 donors per cluster, and tried different algorithms to train the classification, without success. Figure 19 shows results for most of the applied methods, including those we assumed capable of deriving the "ideal" clustering.

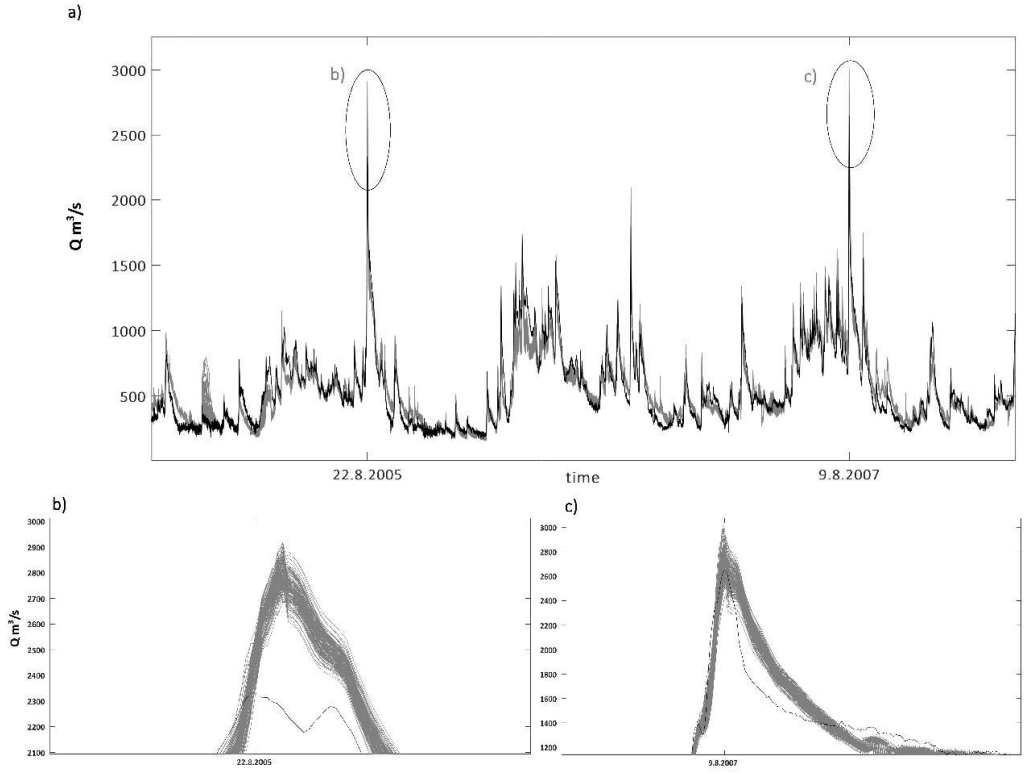


Figure 18 a) Simulated (grey) and observed (black) discharge of the Aare at Untersiggenthal; b) zoom of the peak on the 22nd August of 2005; c) zoom of the peak on the 9th August of 2007.

Following the expert meeting of 24th November 2017, B. Schaefli recommended to pursue a more physical regionalization in the HBV model (similar to that proposed by Bárdossy 2007), as trying to regionalize heuristic ensembles of parameter sets very likely cannot give good results. Furthermore, she suggested that the influence of the ungauged catchments on the discharge maximum at the Aare catchment scale should be tested. This can be tested by assigning any randomly drawn parameter set (from the pool of available parameter sets) to the ungauged catchments and testing the sensitivity of

the peak flows to these choices. If the sensitivity is small, then even a regionalisation method which is not entirely satisfying can be retained.

Hence, for the sub-catchments to be regionalized we picked first 5 parameter sets (evenly distributed between the best and the worst parameter set) from each of the donor sub-catchments, randomly shuffled these, and run simulations for 24 years (between 1.1.1991 and 31.12.2014).

Generally, we found no wide spread or a consistent bias in the simulations, and the spread of the two major floods of 2005 and 2007 varies depending on the flood (Figure 18) and is in the same order of magnitude of the uncertainty resulting from the rating curve uncertainty (see section 4 discharge uncertainty).

Subsequently we tested the regionalization method proposed by Bárdossy (2007), where in a first step mean discharge (MQ) and standard deviation of discharge (SQ) are regionalized, and then parameter sets are selected that are able to reproduce MQ and SQ. This method conveys the importance of selecting donors that have similar water balance and reactivity as the receiver – both characteristics which are not necessarily reflected in the hydrological similarity, and not explicitly measured by this- and it allows to consider parameter sets coming from any other gauged sub-catchment and test it on any other sub-catchment, increasing on one side the pool of considered parameter sets, and on the other side the pool of sub-catchments on which the method can be tested.

Also here we tried different regionalization methods for MQ and SQ, fitting linear and potential functions, and performing a stepwise model selection by the Akaike information criterion, as well as applying a random forest regression. We compared the different methods with already existing datasets, such as the gridded mean annual runoff made available from the CCHydro project for the present day (1980-2009) by the FOEN, the mean annual runoff for the river network data set MQ-GWN-CH for the reference period of 1981-2000 by the FOEN and finally with data available from Plate 5.6 *Flood Discharge- Analysis of long-standing measurement series* of the Hydrological Atlas of Switzerland (in the latter both MQ and SQ were available for some sub-catchments). We found the random forest regression to have the best overall performance, with maximum deviations of ±20% from the observations for the period 1981-2014. Comparing our results with the CCHydro data set show a fairly good agreement (median of deviation 12%), and comparing the two datasets by the FOEN resulted in generally significantly higher differences.

We modified the definition of good parameter vectors defining those as the ones with model efficiencies Nash-Sutcliffe (NSE), Kling-Gupta (KGE) and the weighted multi-objective function >= 0.5, ending up with 1772 parameter sets as potential candidates to be transferred to another sub-catchment. We defined a parameter set as a reasonable candidate and transferred it in case the absolute deviation of MQ and SQ simulated didn’t exceed 30% from the regionalized MQ and SQ. We stress again that according to this procedure, only those parameter sets that provide a reasonable water balance and variability of discharge for the target ungauged sub-catchment are considered for transfer (Bárdossy, 2007).

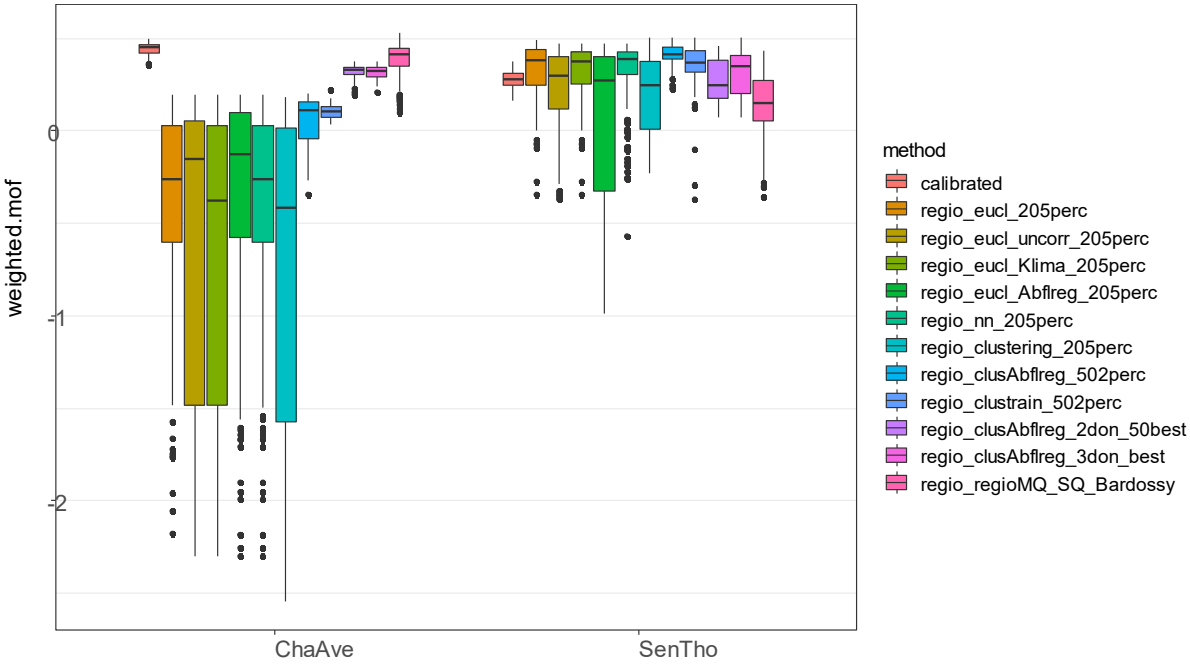


Figure 19 Simulated weighted multi-objective of different tested regionalization methods for the catchments Chandon-Avenges (ChaAve) and Sense-Thörishaus (SenTho). regio_eucl_205perc: Euclidean distance from 5 donor catchments with 20 parameter sets each selected uniformly between worst and best set; regio_eucl _uncorr_205perc: uncorrelated Euclidean distance, from 5 donor catchments and 20 parameter sets each; regio_eucl_Klima_205perc: Euclidean distance with climatic regions, from 5 catchments and 20 parameter sets each; regio_eucl_Abflreg_205perc: Euclidean distance with discharge regimes, from 5 catchments and 20 parameter sets each; regio_nn_205perc: Nearest Neighbours, from 5 catchments and 20 parameter sets each; regio_clustering_205perc: Clustering with k-means, from 5 catchments and 20 parameter sets each; regio_clusAbflreg_502perc: Clustering with discharge regime, from 2 donor catchments with 50 parameter sets each selected uniformly between worst and best set; regio_clustrain_502perc: „Ideal" clustering, from 2 donor catchments 50 parameter sets each; regio_clusAbflreg_2don_50best: Clustering with discharge regimes, from 2 donor catchments 50 best parameter sets each; regio_clusAbflreg_3don_best: Clustering with discharge regimes, from 2 donor catchments for each 33 best parameter sets and from a third donor catchment 34 best parameter sets; regio_regioMQ_SQ_Bardossy: regionalization after Bárdossy (2007).

Generally, the application of this method is beneficial, especially better KGE values were found compared to the other methods. In very few cases –mainly for small sub-catchments located around the Jura lakes- it was even able to outperform the calibration (see results for the sub-catchment ChaAve (Figure 19, left) while for others it resulted in a deterioration of the model performance (see results for the sub-catchment SenTho (Figure 19, right). While for some sub-catchments a large number of parameter sets was selected (e.g. for SarBro more than half of the 1772 parameter sets was transferred), for some ungauged sub-catchments, not a single parameter set was selected from the 1772 parameter sets. While this method seems promising, it would still need more analysis, fine-tuning and an eventual extension, what would have caused a time schedule issue in the project. Therefore and considering the results of the sensitivity analysis, we finalized the regionalization by choosing the version producing the overall best results: clustering using discharge regime as discriminant, with 2 donors selecting the 50 best parameter sets.

3.3 Calibration and plausibility checks of the flood routing model RS Minerve with the hydraulic 2D model BASEMENT

3.3.1 Calibration data from the hydraulic 2D model

The flood routing in RS Minerve was calibrated with results from the hydraulic 2D model BASEMENT v3. From BASEMENT, the geometry data (width, slope angle, bank height and roughness coefficient)

were read out section by section and transferred to RS Minerve. In addition, the capacities and flow times resulting from the 2D hydraulic simulations were transferred to RS Minerve. The different retention areas along the channel were determined from the flood areas of the simulations. The retention areas were implemented in RS Minerve by means of artificial, very wide and slowly flowing channels.

With the basic data from the hydraulic model, the interactions of tributaries and retention areas was calibrated based on the flood event 2005 and validated with the flood event 2007.

3.3.2 Plausibility checks for very large events

Since simulated hydrological hydrographs are of interest for events with much higher discharges than those measured in 2005 and 2007, RS Minerve also had to represent the interactions of floods for events that activate much larger flood areas than contained in the data.

For such high discharges the event simulation with RS Minerve was compared with the calculation of the floods with BASEMENT. The comparison was carried out under the assumption that BASEMENT models the flow processes in the surrounding area realistically.

Characteristic scenarios for comparative calculation

For the comparison between the two models RS Minerve and BASEMENT, a total of six scenarios were selected. Whereby a range of outflows as wide as possible at the Stilli transfer point (SSHSSI) was chosen as the selection criterion. The following scenarios were selected from the extreme value analysis (see Figure 20):

- Peak-1-scen-96: Top Event of the 289’000-year long simulation
- Peak-1-scen-38: Second major event in the order of FL4 (probability roughly 10^{-4} in Stilli)
- FL4-med-scen-215: Scenario in the order of magnitude of a FL4 in Stilli
- FL4-med-scen-129: Scenario in the order of magnitude of a FL4 in Stilli
- FL3-med-scen-44: Scenario in the order of FL3 (probability roughly 10^{-3} in Stilli)
- FL3-med-scen-232: Scenario in the order of magnitude of a FL3 in Stilli

The hydrographs at the various transfer points in the total system along the Aare River and the main tributaries of the Saane, Emme, Reuss and Limmat Rivers were compared. The shape of the hydrographs was visually assessed and deviations in peak discharge values were determined. The most important transfer points are those near the assessment sites:

- Assessment site Mühleberg: Aare Halen (SSASSB) and Saane Gümmenen (SSJSSB)
- Assessment site Olten and assessment site BP Gösgen: Aare Aarburg (SSESSF)
- Assessment site PSI and assessment site Beznau: Aare Stilli (SSHSSI)

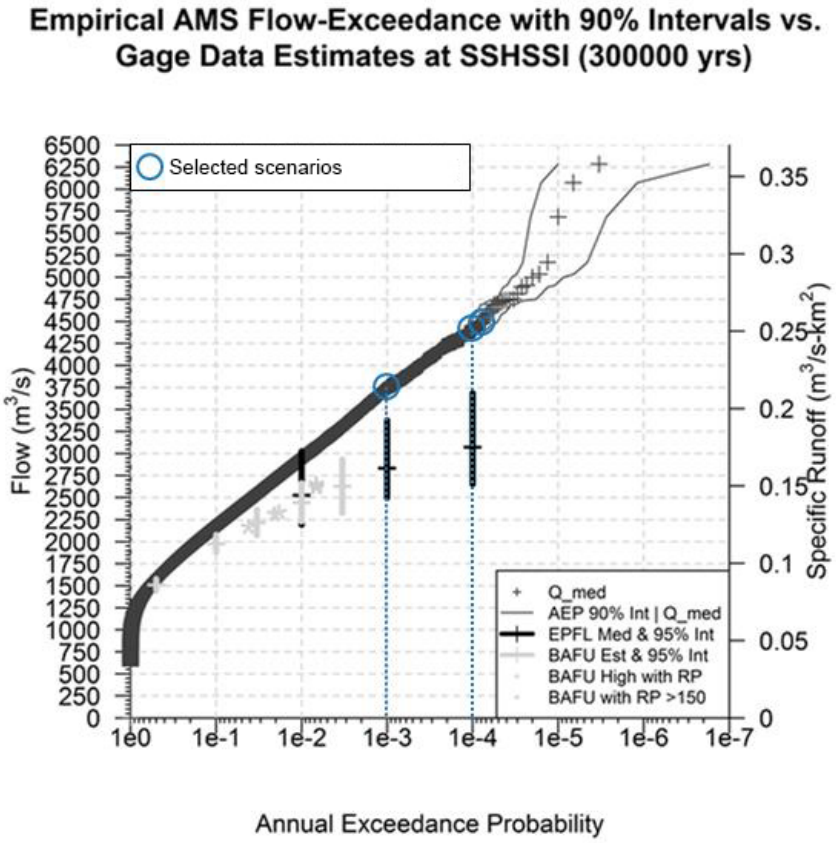
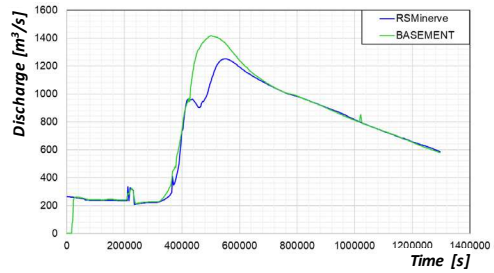


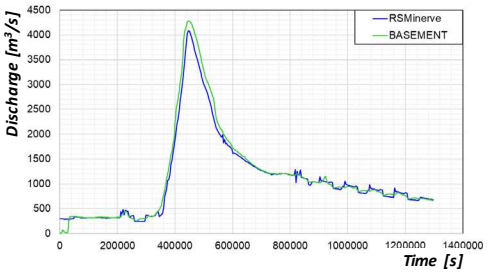
Figure 20 Extreme value analysis at Transfer point Stilli (SSHSSI) with indication of the selected hydrological scenarios for the comparison of RS Minerve and BASEMENT (marked in blue).

Hydrograph deviations at different transfer points

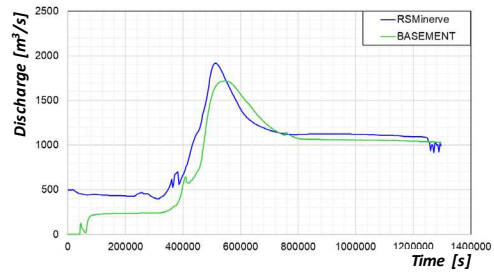
Overall, the comparison shows that the discharge hydrographs of the two models have a good to very good fit. The hydrographs above Lake Neuchâtel, Lake Biel and Lake Murten (SSASSB, SSBSSC, SSBSSJ) and the hydrographs of the remaining tributaries (SSKSSD, SSLSSH, SSMSSH) are almost identical.



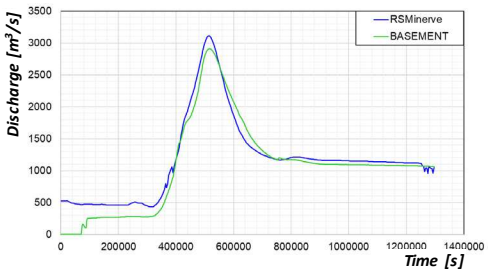
Transfer point Halen (SSASSB)



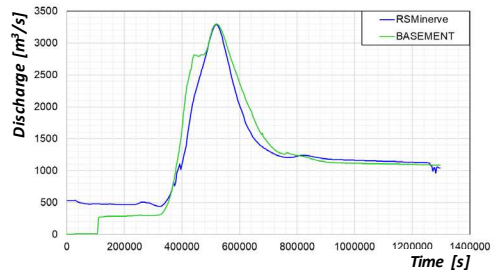
Transfer point Golaten (SSBSSC)



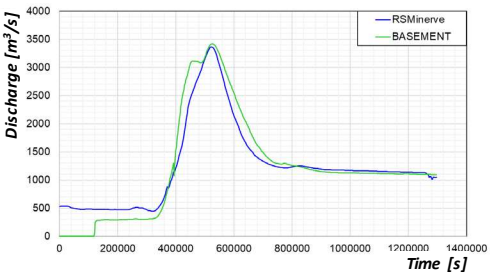
Transfer point Solothurn (SSCSSD)



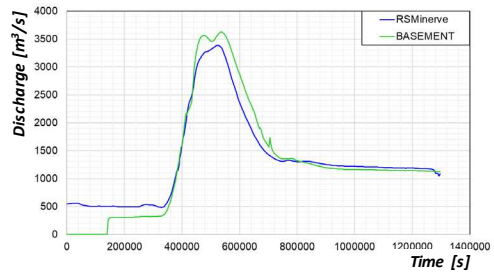
Transfer point Walliswil (SSDSSE)



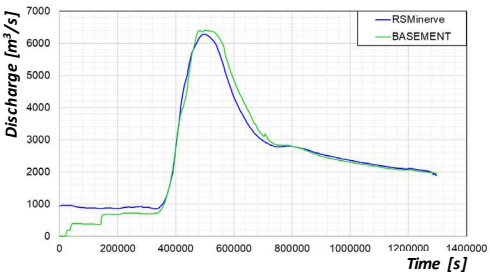
Transfer point Aarburg (SSESSF)



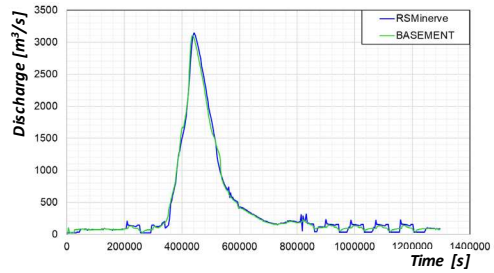
Transfer point Aarau (SSFSSG)



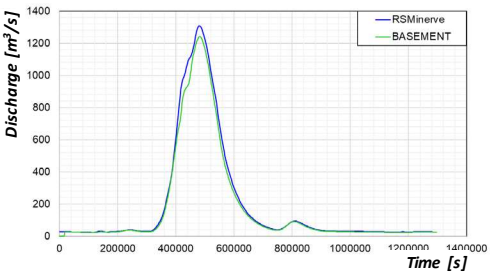
Transfer point Brugg (SSGSSH)



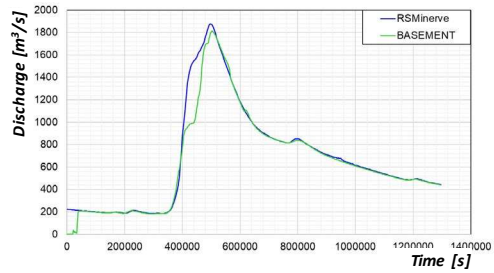
Transfer point Stilli (SSHSSI)



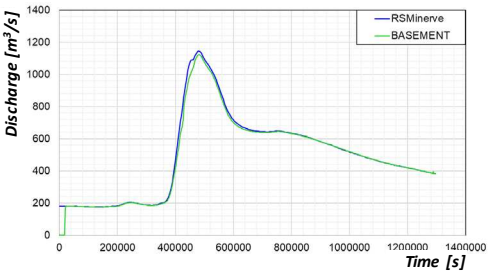
Transfer point Gümmenen (SSJSSB)



Transfer point Derendingen (SSKSSD)

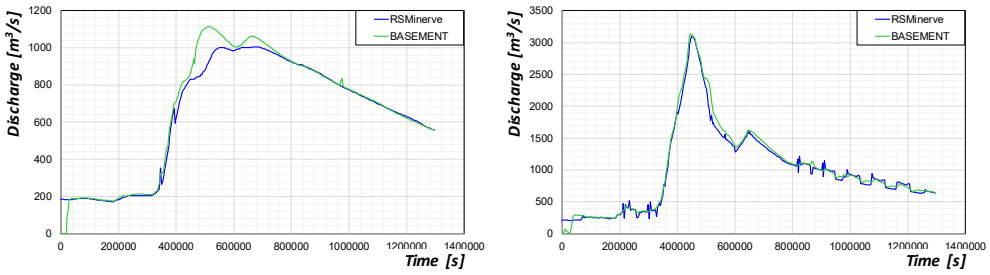


Transfer point Windisch (SSLSSH)



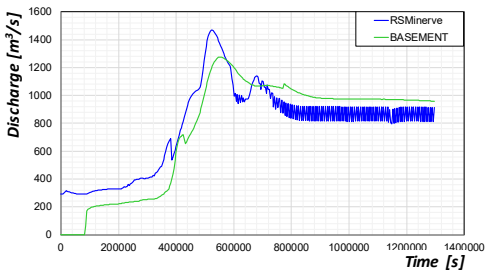
Transfer point Turgi (SSMSSH)

Figure 21 Comparison of scenario Peak-1-scen-96.

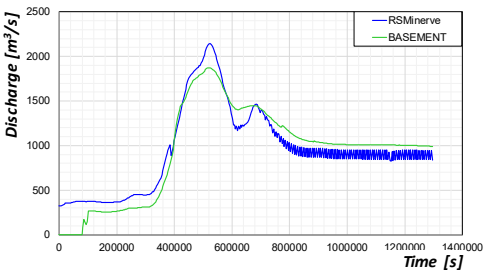


Transfer point Halen (SSASSB)

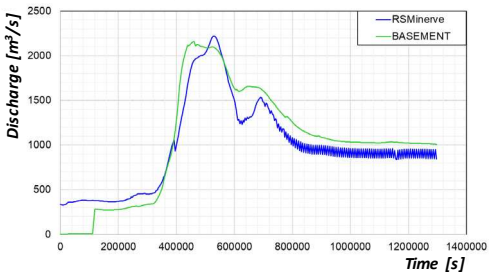
Transfer point Golaten (SSBSSC)



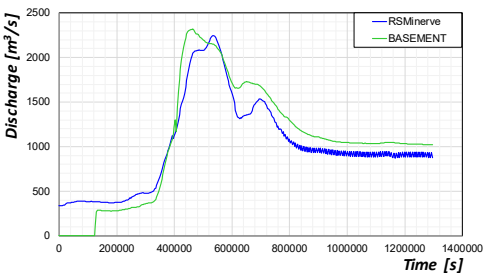
Transfer point Solothurn (SSCSSD)



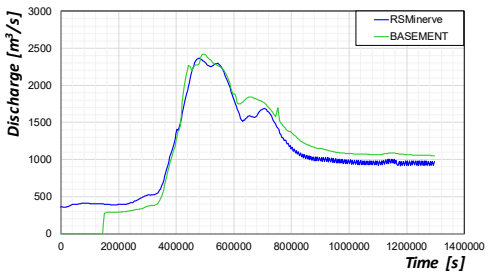
Transfer point Walliswil (SSDSSE)



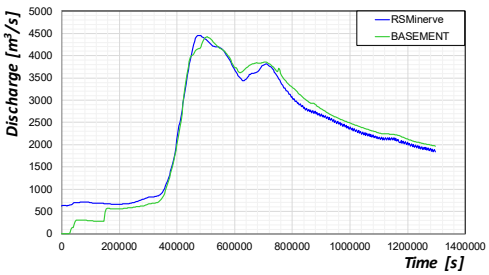
Transfer point Aarburg (SSESSF)



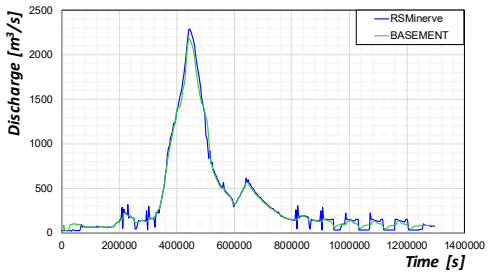
Transfer point Aarau (SSFSSG)



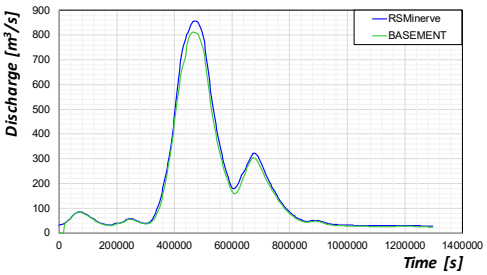
Transfer point Brugg (SSGSSH)



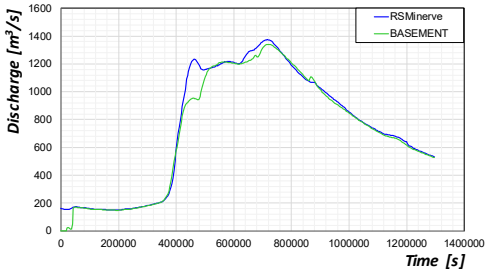
Transfer point Stilli (SSHSSI)



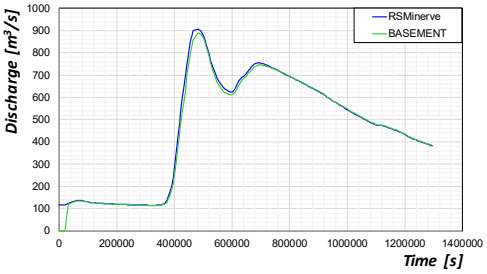
Transfer point Gümmenen (SSJSSB)



Transfer point Derendingen (SSKSSD)

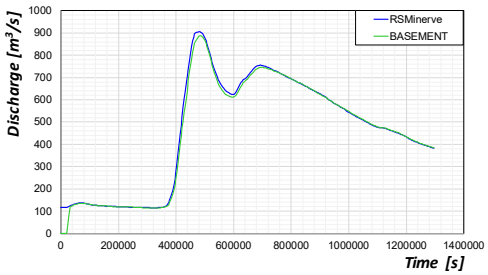


Transfer point Windisch (SSLSSH)

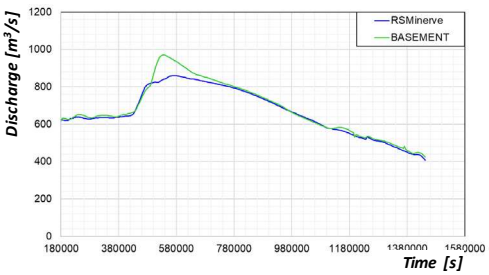


Transfer point Turgi (SSMSSH)

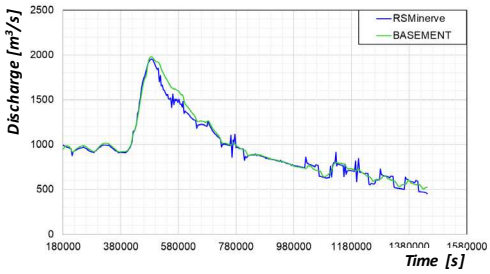
Figure 22 Comparison of scenario Peak-1-scen-38.



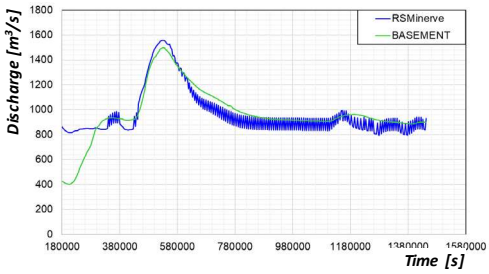
Transfer point Halen (SSASSB)



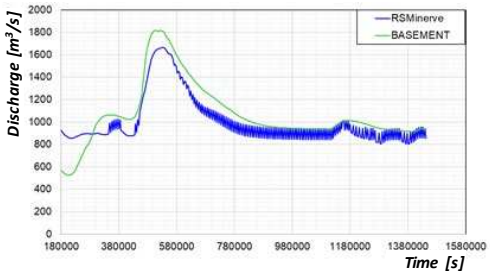
Transfer point Golaten (SSBSSC)



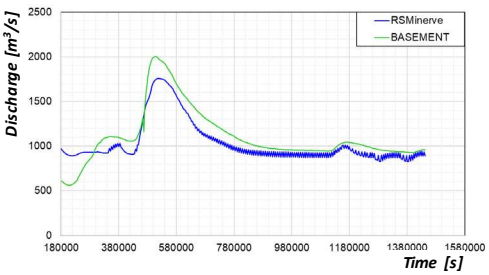
Transfer point Solothurn (SSCSSD)



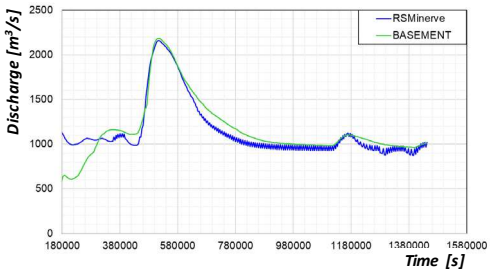
Transfer point Walliswil (SSDSSE)



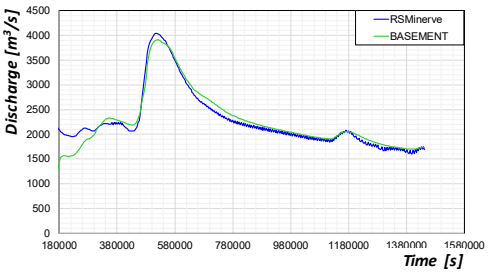
Transfer point Aarburg (SSESSF)



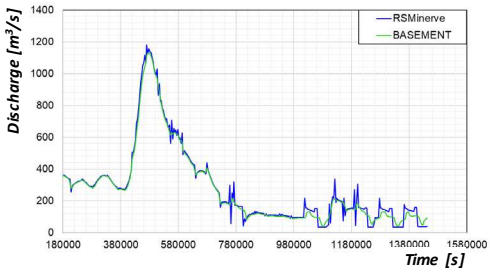
Transfer point Aarau (SSFSSG)



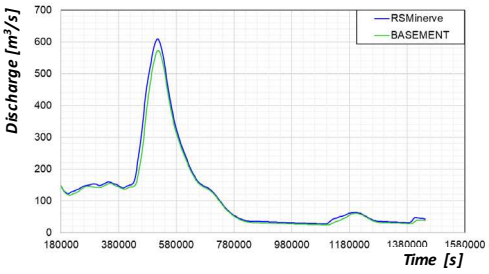
Transfer point Brugg (SSGSSH)



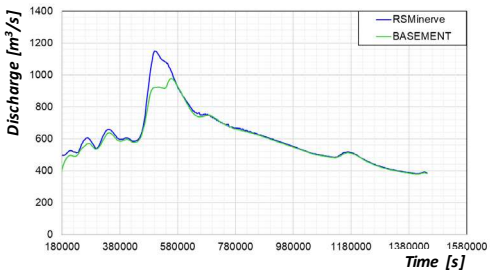
Transfer point Stilli (SSHSSI)



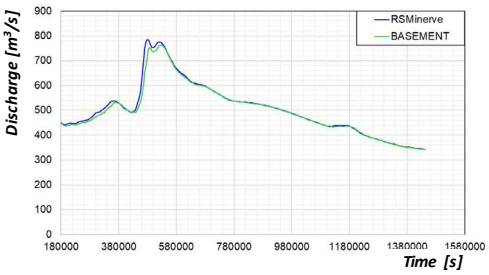
Transfer point Gümmenen (SSJSSB)



Transfer point Derendingen (SSKSSD)

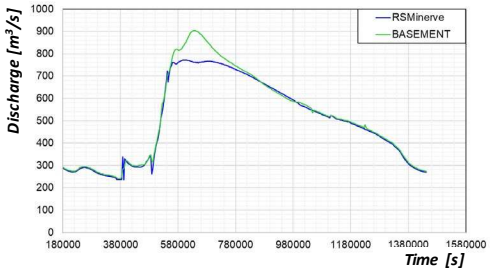


Transfer point Windisch (SSLSSH)

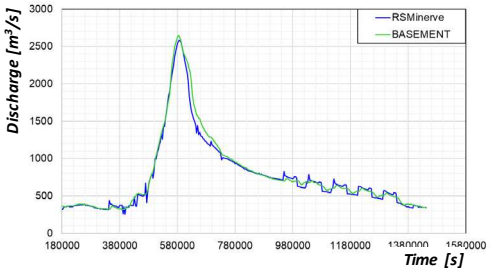


Transfer point Turgi (SSMSSH)

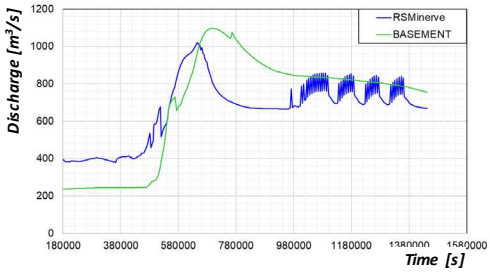
Figure 23 Comparison of scenario FL4-med-scen-215.



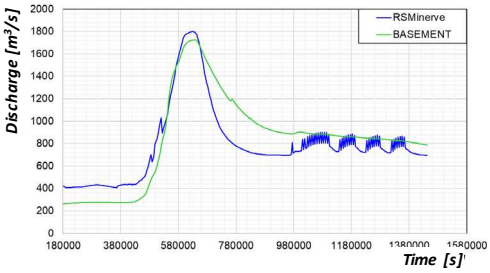
Transfer point Halen (SSASSB)



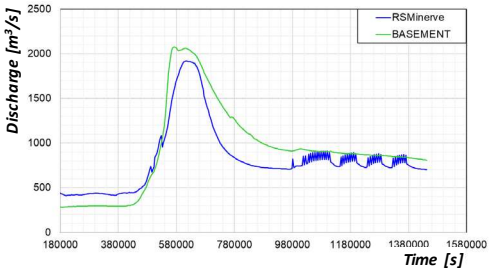
Transfer point Golaten (SSBSSC)



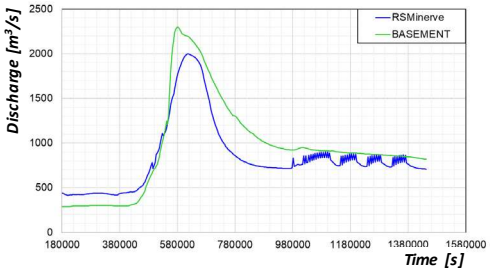
Transfer point Solothurn (SSCSSD)



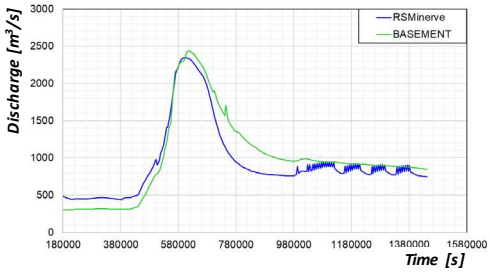
Transfer point Walliswil (SSDSSE)



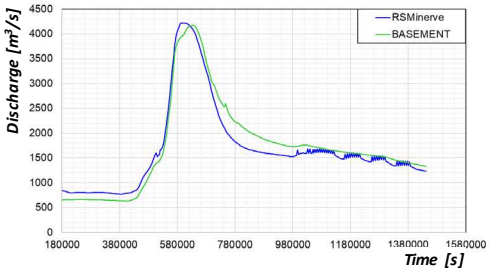
Transfer point Aarburg (SSESSF)



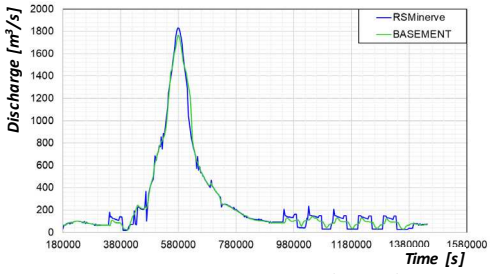
Transfer point Aarau (SSFSSG)



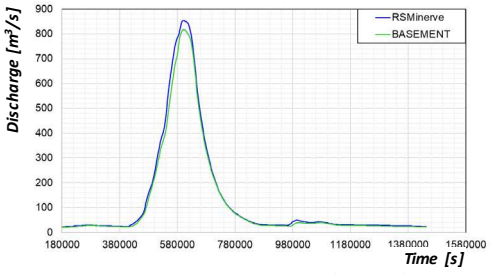
Transfer point Brugg (SSGSSH)



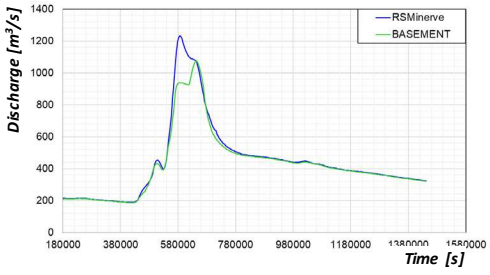
Transfer point Stilli (SSHSSI)



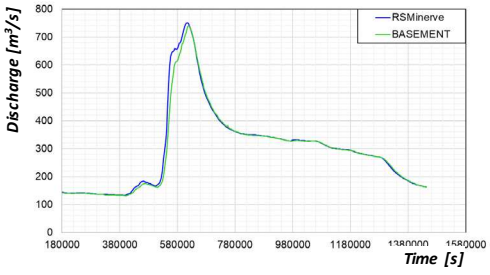
Transfer point Gümmenen (SSJSSB)



Transfer point Derendingen (SSKSSD)

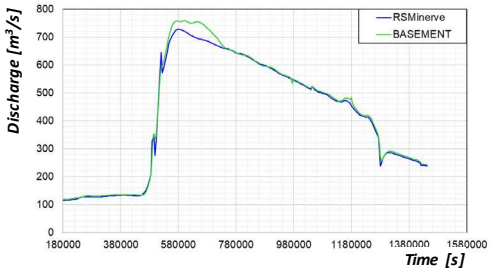


Transfer point Windisch (SSLSSH)

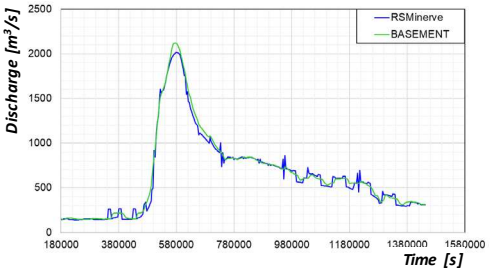


Transfer point Turgi (SSMSSH)

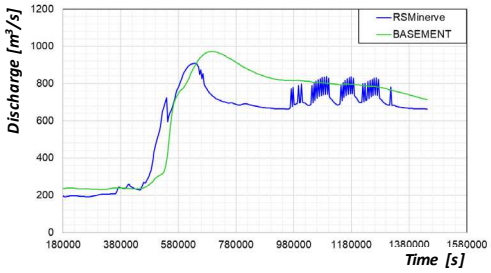
Figure 24 Comparison of scenario FL4-med-scen-129.



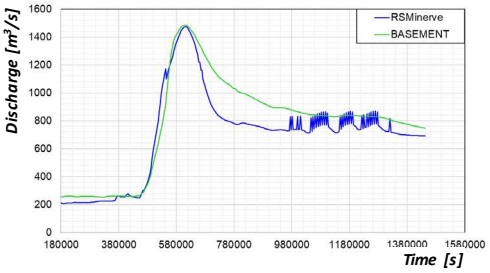
Transfer point Halen (SSASSB)



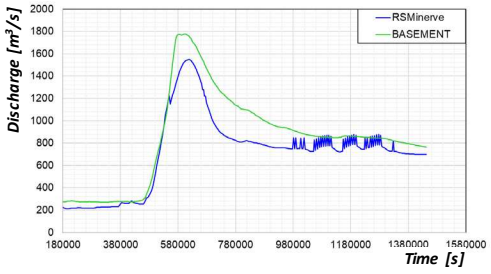
Transfer point Golaten (SSBSSC)



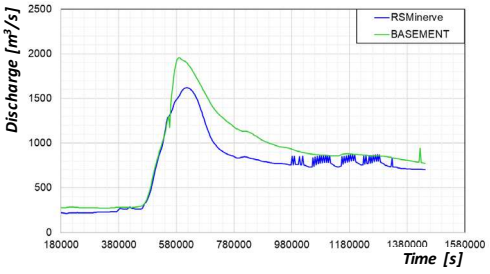
Transfer point Solothurn (SSCSSD)



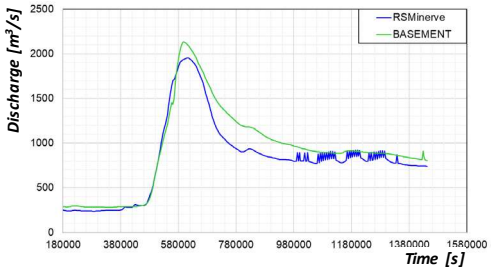
Transfer point Walliswil (SSDSSE)



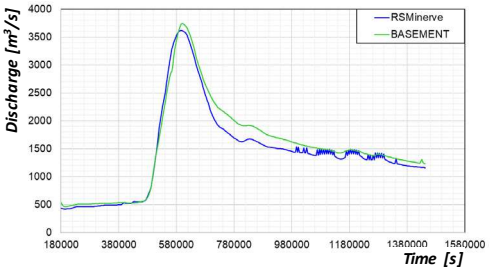
Transfer point Aarburg (SSESSF)



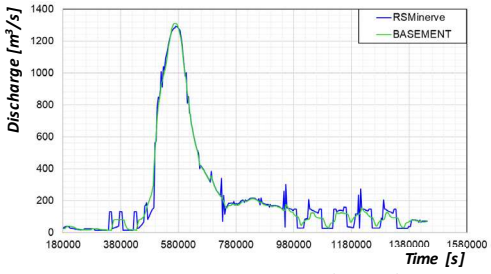
Transfer point Aarau (SSFSSG)



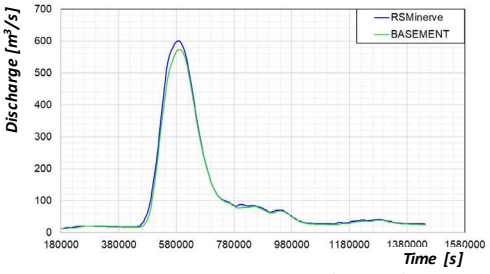
Transfer point Brugg (SSGSSH)



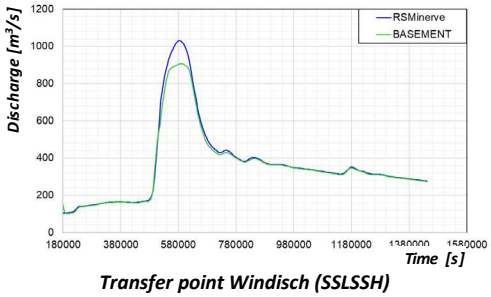
Transfer point Stilli (SSHSSI)



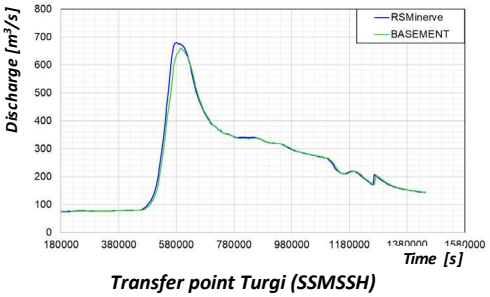
Transfer point Gümmenen (SSJSSB)



Transfer point Derendingen (SSKSSD)

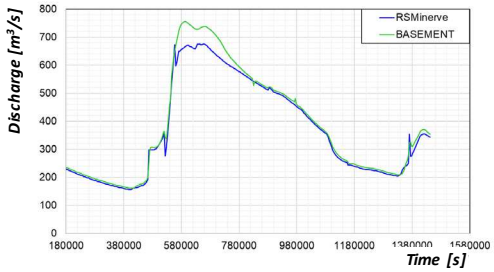


Transfer point Windisch (SSLSSH)

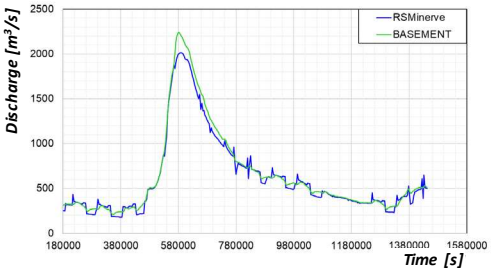


Transfer point Turgi (SSMSSH)

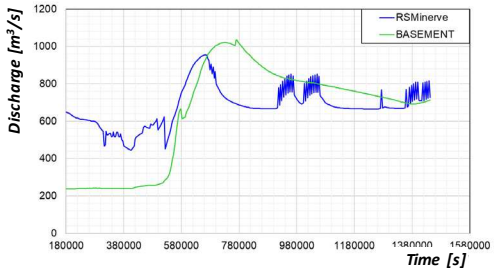
Figure 25 Comparison of scenario FL3-med-scen-44.



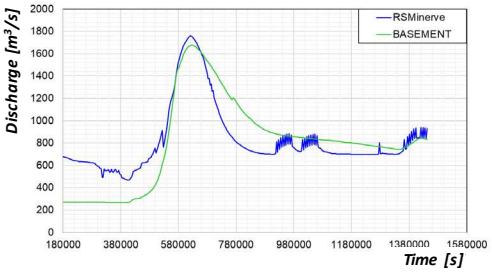
Transfer point Halen (SSASSB)



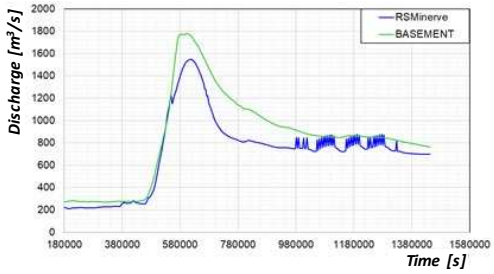
Transfer point Golaten (SSBSSC)



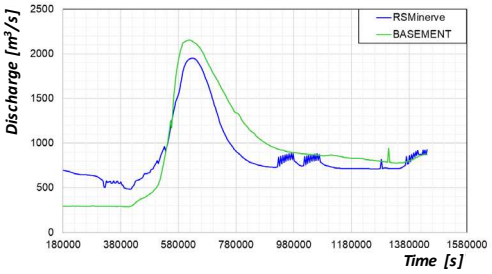
Transfer point Solothurn (SSCSSD)



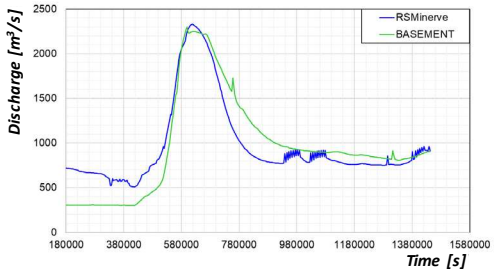
Transfer point Walliswil (SSDSSE)



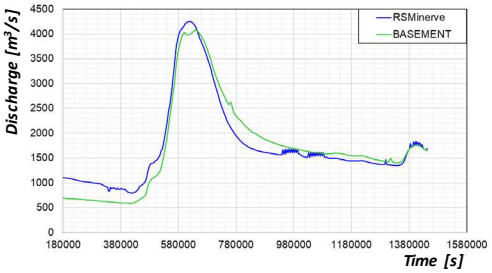
Transfer point Aarburg (SSESSF)



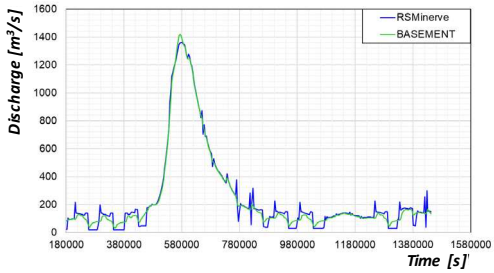
Transfer point Aarau (SSFSSG)



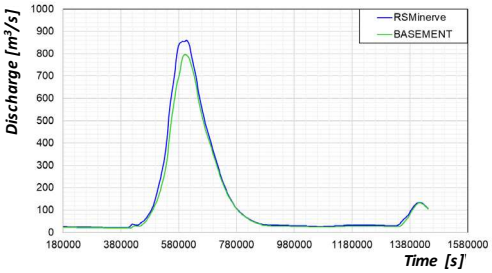
Transfer point Brugg (SSGSSH)



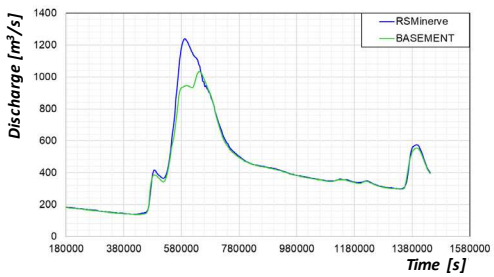
Transfer point Stilli (SSHSSI)



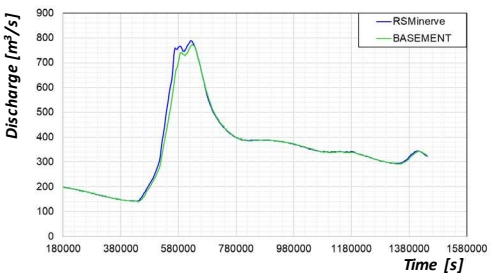
Transfer point Gümmenen (SSJSSB)



Transfer point Derendingen (SSKSSD)



Transfer point Windisch (SSLSSH)



Transfer point Turgi (SSMSSH)

Figure 26 Comparison of scenario FL3-med-scen-232.

Below Lake Neuchâtel, Lake Biel and Lake Murten some differences between the two models are noted:

- At transfer point SSCSSD (Solothurn), the differently implemented regulation of the Port Weir is evident.
    - In BASEMENT, less water flows through the weir at the start of the hydrological scenario, which increases the retention volume.
    - Both models switch from normal to flood mode at about the same time.
    - The maximum discharge through the Port Weir as well as the flow time in the discharge corridor of the Old Aare takes longer in BASEMENT than in RS Minerve in all scenarios. The time lag between the two runoff peaks is usually between 10–12h but can also be as much as 24h (FL4-med-scen-215 and FL3-med-scen-44).
    - The peak discharge achieved at the transfer point SSCSSD (Solothurn) is higher for the largest hydrological events (Peak1-scen-96 and Peak1-scen-38) in RS Minerve than in BASEMENT (200–400 m^3/s). In the remaining scenarios the peak discharge of BASEMENT is about 100–150 m^3/s higher than that of RS Minerve.
- At transfer point SSDSSE (Walliswil), the two hydrographs approach a uniform shape again by the superposition of the flood hydrograph of the Aare River with that of the Emme River.
    - The relatively short and pronounced flood hydrograph of the Emme River leads to a new flood peak in the already strongly attenuated Aare River flood hydrograph. In both models, the inflow of the Emme River has a similar effect on the discharge hydrograph at transfer point SSDSSE.
    - The measured peak discharge at transfer point SSDSSE in RS Minerve is about 100 m^3/s higher in all scenarios than in BASEMENT.
- At transfer point SSESSF (Aarburg), the peak discharge timing is the same for both models in almost all scenarios. In the very large scenarios (Peak1-scen-96 and Peak1-scen-38) peak runoff is also very similar in both models. For the remaining scenarios (FL4 and FL3) the peak runoff of the RS Minerve hydrograph is continuously higher by 100–400 m^3/s.

Below the confluence of the Reuss and Limmat Rivers (transfer point SSHSSI, Stilli), the hydrographs are almost identical in shape, peak discharge value and time of peak discharge.

Changes in flood volume

In addition to the hydrograph, we compared the volume of the flood event (Figure 27Figure 27). To ensure comparability of the volumes and to avoid distortions due to possible initial differences, the hydrographs of the two models were cut to the same range (200’000–1’296’000 s). The change in volume allows statements about the retention volumes in the individual subsystems, since the input hydrographs in both models are the same (inflows to the POI + diffuse inflows).

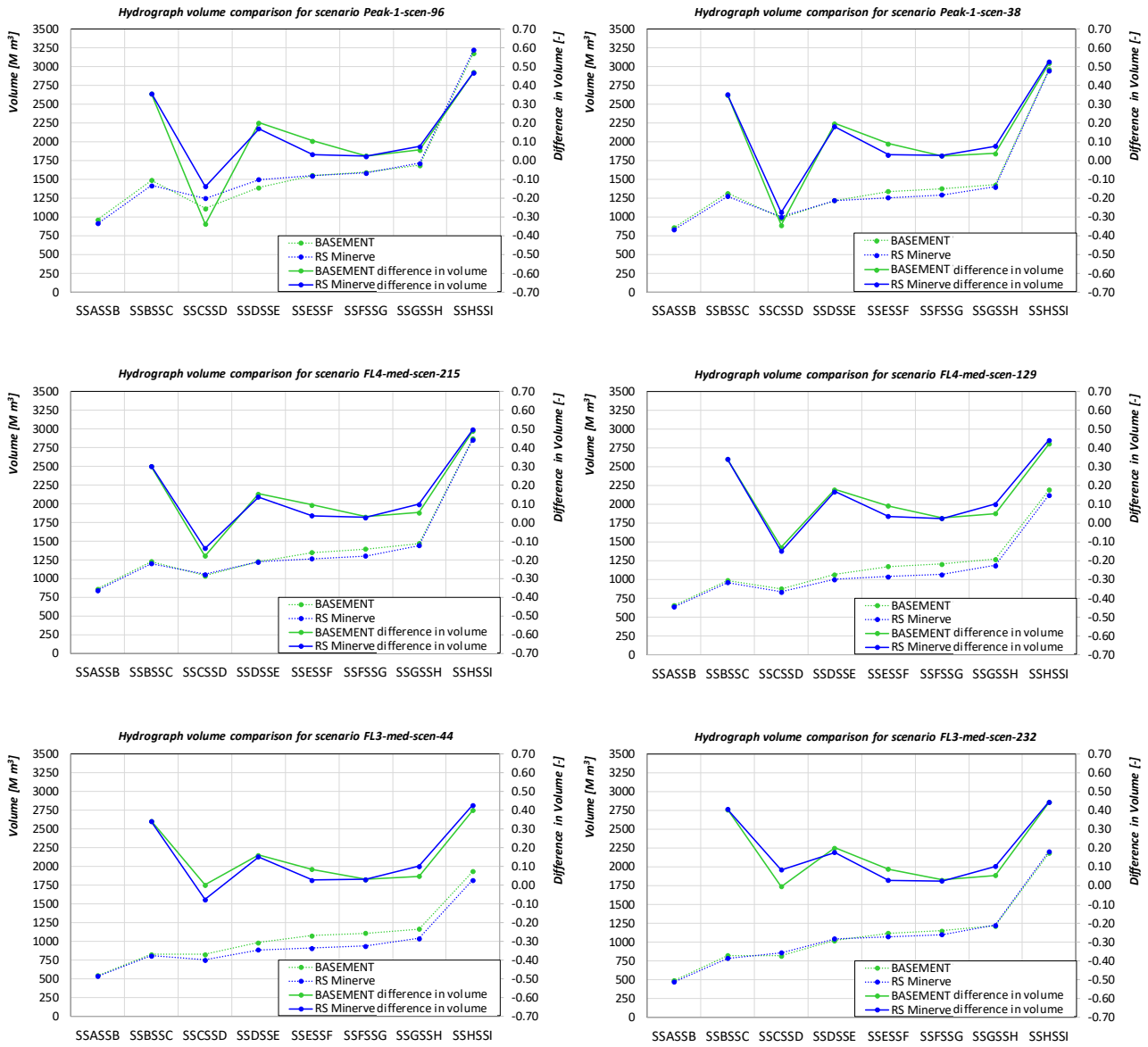


Figure 27 Comparison of the hydrograph volumes for the selected hydrological scenarios.

The change in volume per subsystem shows that the two models represent the retention in a similar way. For example, the volume curves in subsystem B shift parallel to each other, which indicates that flood volumes and retention behave identically in this subsystem.

In subsystem C the retention in BASEMENT is larger than in RS Minerve in most scenarios, but in subsystems D and E the retention in RS Minerve is slightly larger.

In all scenarios, the difference in retention volume in subsystem E behaves more or less similar. In subsystem F, the retention is modeled similarly in both models (the volume curves shift parallel to each other). In subsystem G, the retention in BASEMENT is somewhat larger than in RS Minerve, which is compensated by a reverse behavior in subsystem H. In subsystem G, the retention in BASEMENT is slightly larger than in RS Minerve.

Overall, the volume curves and the relative changes are similar in both models. There are subsystems in which the retention in BASEMENT is greater, but also subsystems in which RS Minerve shows a stronger attenuation. These two effects balance each other out over the entire system.

3.3.3 Conclusions

Overall, the model results from RS Minerve can be used at the important transfer points Halen (SSASSB), Gümmenen (SSBSSJ), Aarburg (SSESSF) and Stilli (SSHSSI) without adjustments for extreme value statistics and for modelling in risk assessment.

3.4 Seasonality of flood events

For both precipitation and discharge, the seasonality of the simulated events was evaluated by comparing the pseudo-observations to the GWEX simulations. In addition, the seasonality of precipitation events was compared to the seasonality of discharge events.

The seasonality the GWEX simulations and pseudo-observations were computed for precipitation events and peak flow events using an annual maximum approach. Thus, for each simulation year a maximum value of runoff (AMF, mm/h) and a maximum value of precipitation (mm/h) were extracted and the respective date of occurrence was saved. To derive the event discharge, the event discharge sums of 72 and 144 h were extracted and the maximum value for every year and every catchment selected. Note that in this approach the maximum precipitation event might occur at a different time than the maximum runoff event. The time of occurrence of maximum precipitation and maximum peak flow may differ since precipitation is not necessarily the only factor triggering floods in the catchment but there might be large snow melt contributions that lead to large events. Note further that these analyses are done for each catchment independently and thus there is no spatial consistency between different catchments, meaning that in each sub-catchment a different event may be classified as the annual maximum event. The extracted series of annual precipitation maxima and AMFs were then analysed for their seasonality and magnitudes. As the three parameter clusters affect only the magnitude of the peak floods but not their seasonality or the time of the occurrence, the seasonality was analysed using only the median cluster.

3.4.1 Comparison of discharge events from pseudo-observations and GWEX simulations

The seasonality of the simulations with pseudo-observations was similar to the seasonality of the GWEX simulations for most of the catchments (Figure 28), which is apparent from the distribution of the AMFs per catchment in the year, meaning in which month how many AMFs occurred.

The differences between the seasonality of the GWEX simulations and the pseudo observations that are apparent in some catchments should not be over-interpreted, since the sample sizes of GWEX (n = 289000) and the pseudo-observations (n = 85) are very different. This means that we could find events sampled form the longer GWEX series that could have also occurred in the shorter pseudo-observation series but were not observed in the short time of the observations given. As a consequence of the different sample sizes, the seasonality of GWEX simulations and pseudo-observations cannot be compared one to one. Rather, the tendency of the seasonality has to be assessed.

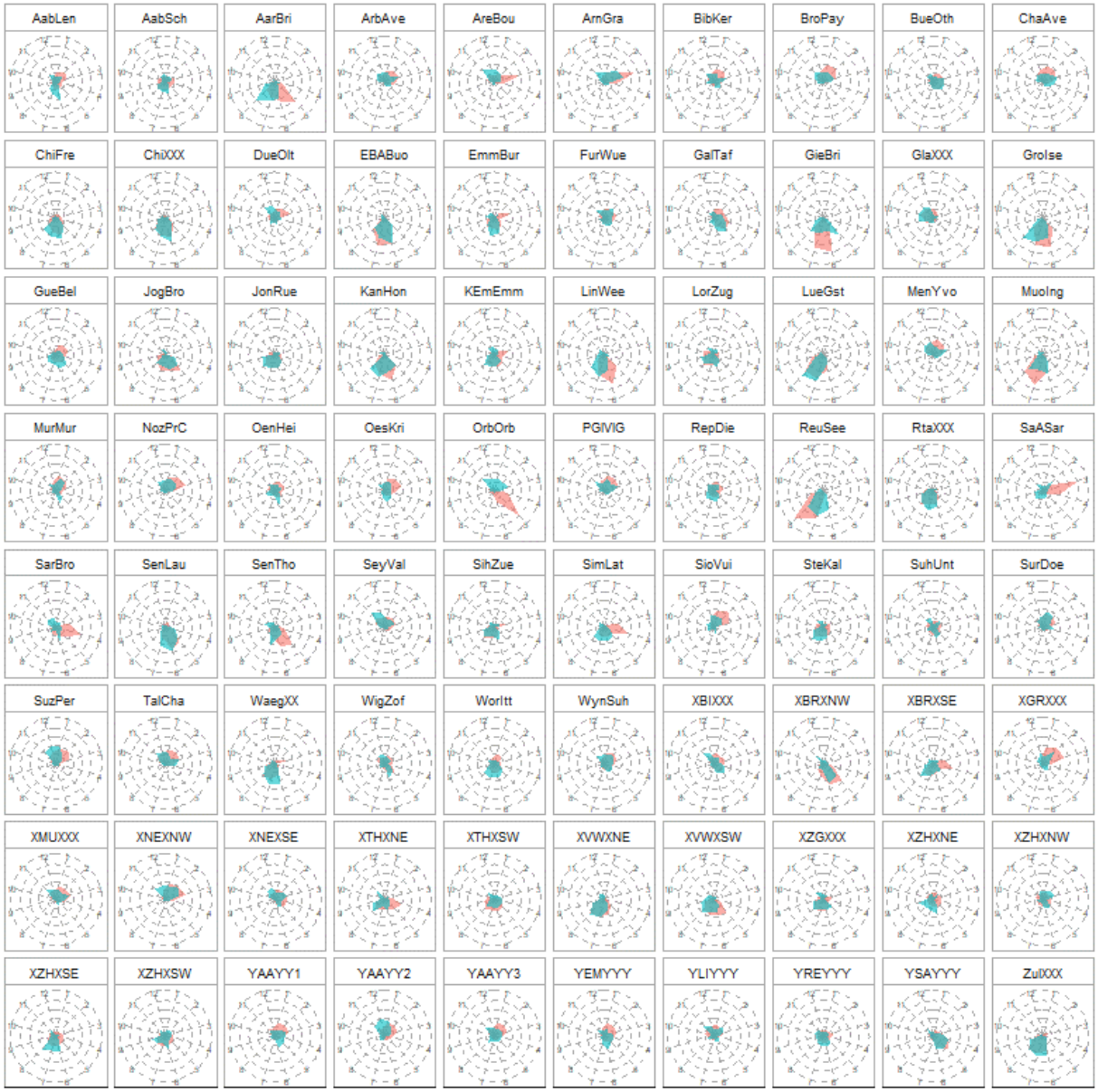


Figure 28 Seasonality of the largest 72 h annual discharge events of the GWEX-simulations compared to the pseudo-observations. The seasonality of the pseudo-observations is shown in turquoise, the seasonality of the GWEX simulations in red.

For the total system we looked at the month of occurrence of the AMF, at the mean date of occurrence across all years (as day of the year DoY) and at the strength of the seasonality (it varies between 0 and 1 – the higher, the stronger the seasonality). All TPs generally show a rather nice and regular distribution, which is either positively skewed or almost Gaussian for most TPs, and mainly concentrated around the spring and summer months (see Figure 29). The strength of seasonality is indeed medium to high (above 0.5 for most TPs), besides for the Lorze (LorReu) and the Emme (SSKSSD). The mean date of occurrence varies between early June (e.g. outlet Lake Biel) and mid-August (e.g. outlet Lorze).

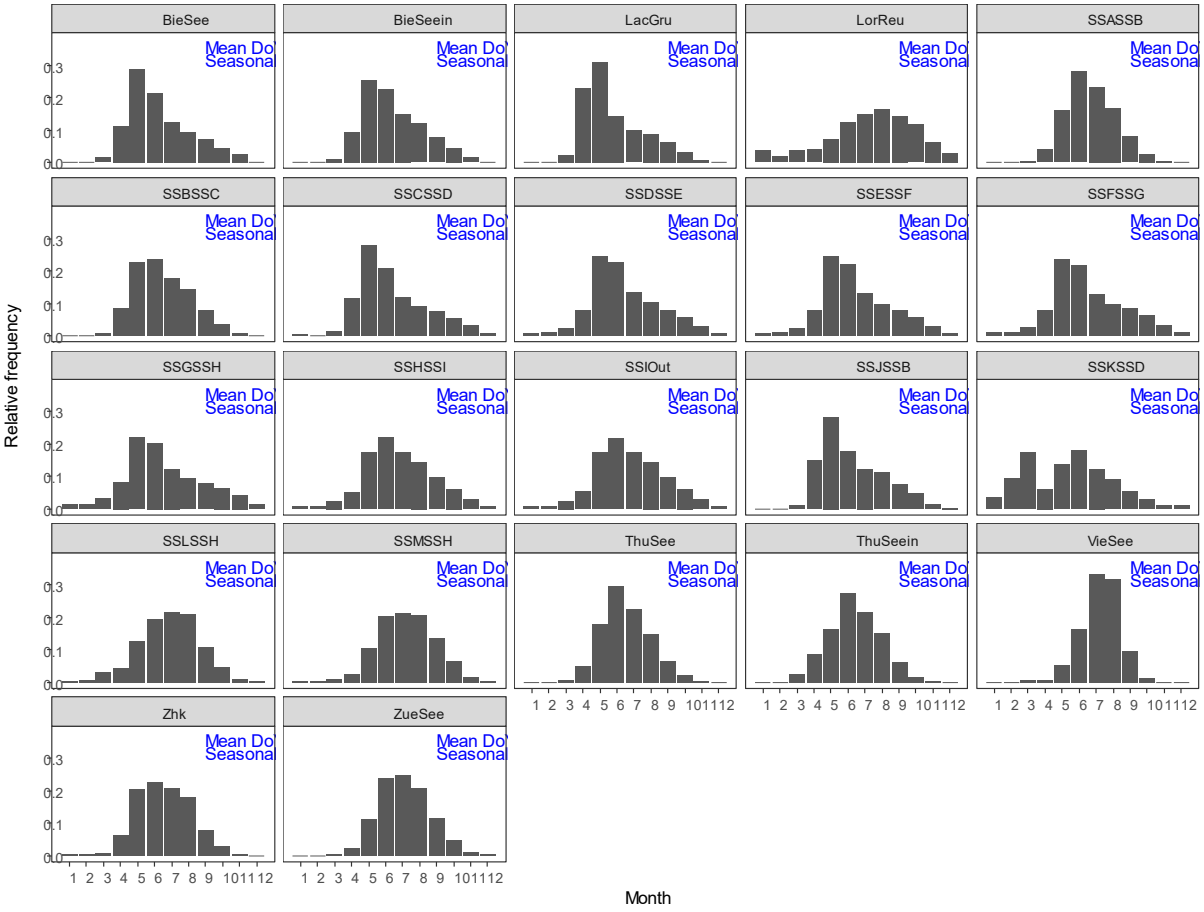


Figure 29 Histograms of the month of AMF occurrence of the GWEX 289'000 simulated years. Numbers in blue are circular summary statistics indicating the mean date of occurrence across all years and the strength of seasonality.

By doing the same kind of analysis for the analysis for the pseudo-observations we can see that the histograms are more spread out, with more events occurring also in winter for some TPs (see Figure 30). The mean difference in seasonality between both data sets for a given TP measures 0.17, with a maximum difference of 0.37. For GWEX we find a slightly earlier mean date of AMF occurrence (except for VieSee), with the mean difference of -12 days, and a maximum difference of -33 days. The distribution resulting at the TPs is resulting from a mix of different regimes and largely affected by the lake regulations, however it can be said that the main imprint and characteristics found in the pseudo-observations are reflected in the GWEX simulations.

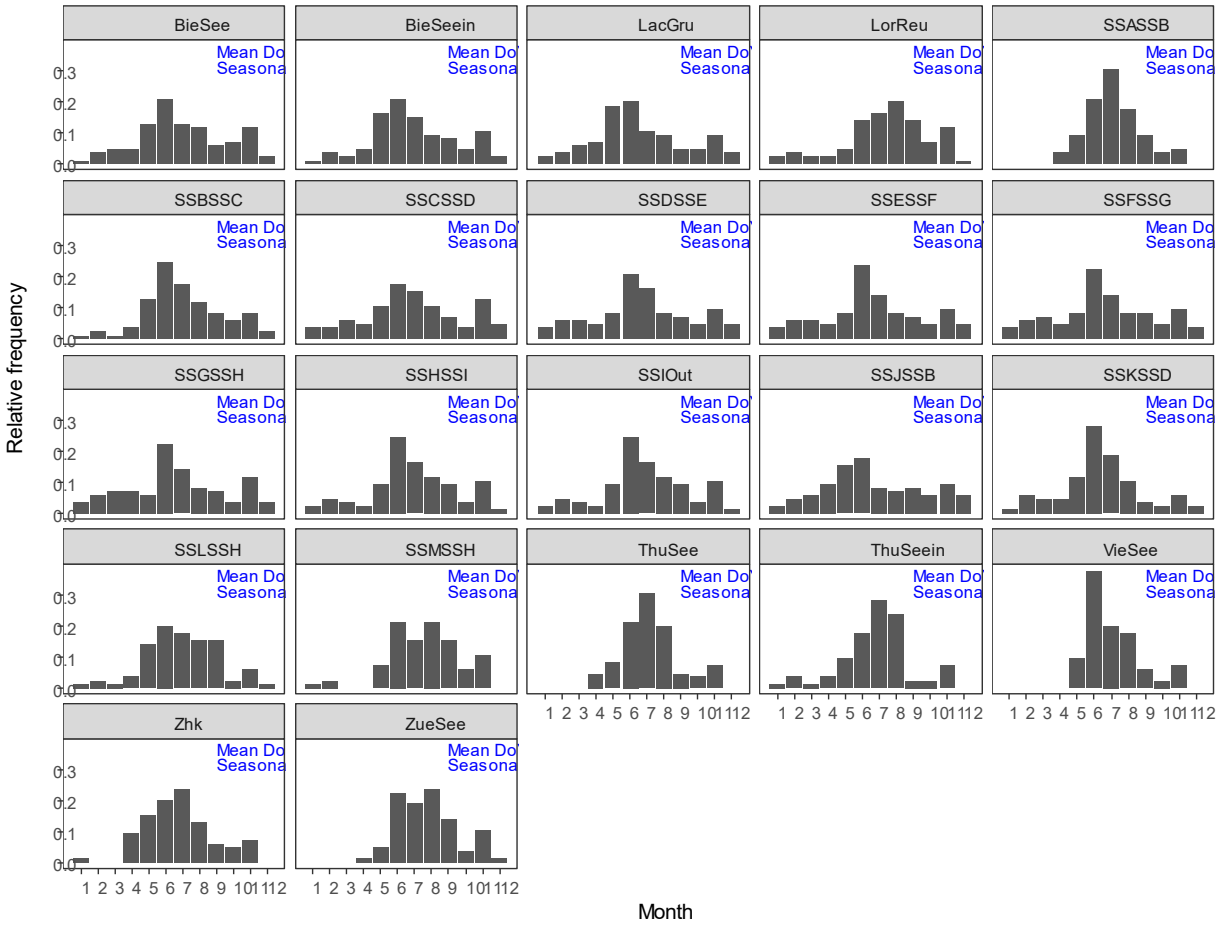


Figure 30 Histograms of the month of AMF occurrence of simulations based on 85 years of pseudo-observations. Numbers in blue are circular summary statistics indicating the mean date of occurrence across all years and the strength of seasonality.

3.4.2 Comparison between precipitation and discharge events

For most catchments, the spatial pattern of the seasonality corresponded to the flood regime of the catchment (Figure 31). The probability of an event to occur was taken from the 289 000 years of simulations of GWEX and counted in which month of the year the annual maximum discharge occurred. For the glacio-nival catchments that are influenced by glacier melt as well as snowmelt there is a clear pattern that most AMFs occur in June, July and August, while the maximum annual precipitation is less pronounced but also occurs during these summer months. For the nival catchments, for which snow melt is dominating the streamflow regime, we also found that peak discharges mainly occurred during the summer months, however starting in May already and stretching to September, i.e. with less pronounced seasonality. Regarding the precipitation in these catchments, here the maximum annual precipitation has a pronounced seasonality with the most frequent occurrence in the months of June, July and August. For the pluvial catchments there is no clear seasonality for the peak discharges.

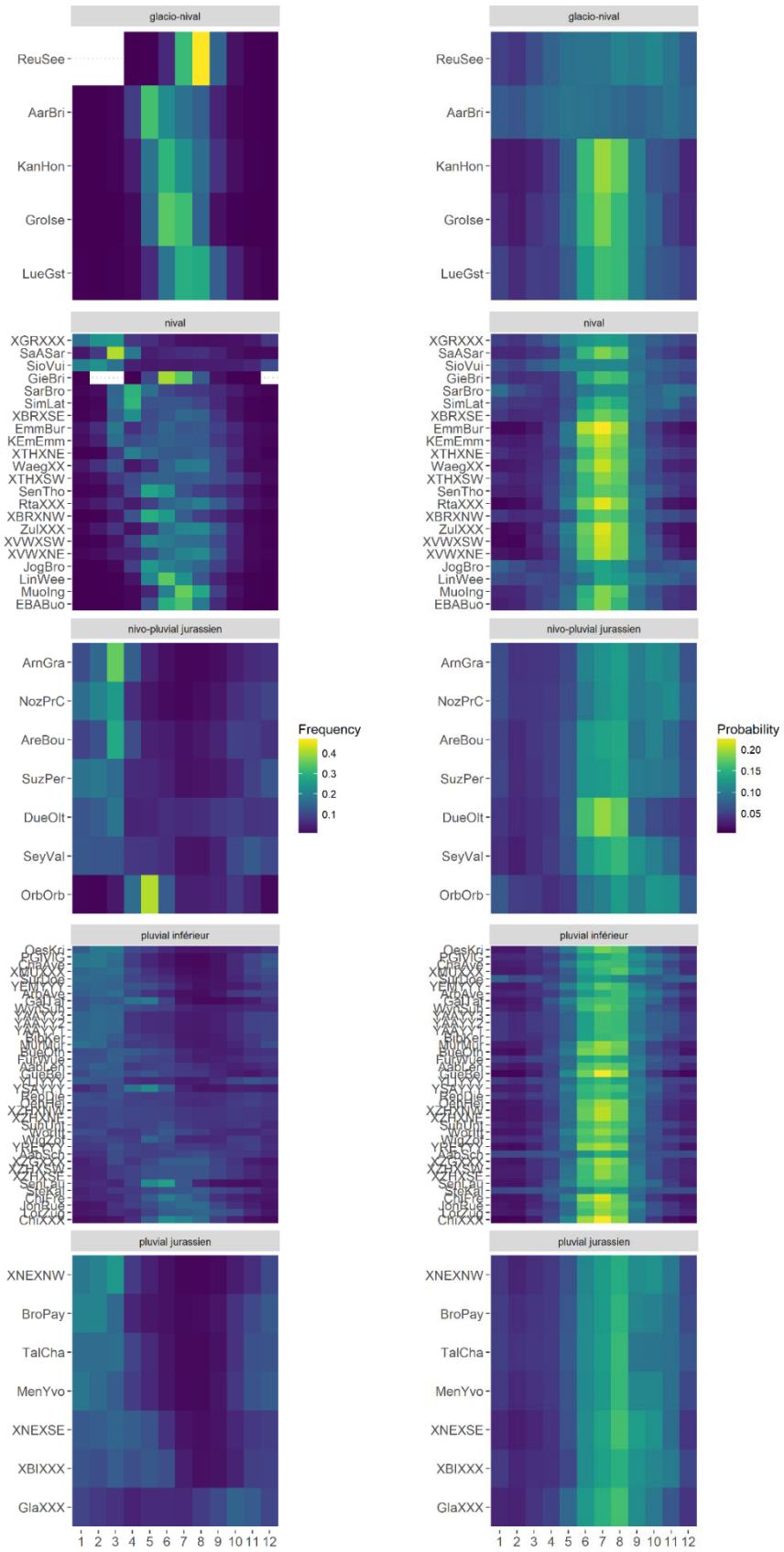


Figure 31 Frequency of occurrence of peak events for each month for the GWEX simulations discharge on the left panel, precipitation on the right panel.

The main difference between the different sub-types of streamflow regimes is however in the occurrence of maximum annual precipitation, where the seasonality is relatively distinct for the pluvial-inférieur regime with events occurring in June, July and August. In the pluvial jurassien regime type,

the highest precipitation events are more stretched over June to September and occur up to November.

3.4.3 Conclusion

From the comparison of the seasonality between pseudo-observations and GWEX simulations, while accounting for the different sample size, we can conclude that the GWEX seasonality is similar to the one we find in the observations.

The seasonality of the highest precipitation and the highest discharge events is similar for the snow and glacier influenced catchments, indicating that snowmelt and precipitation both contribute to the highest discharges. Snow could occur the month preceding the highest discharge events and increase the antecedent wetness of the catchments. For the pluvial catchments, the seasonality of the largest precipitation events is surprisingly large compared to that of the largest discharge events, with the discharge events occurring at the beginning and the end of a calendar year, while the precipitation events are in summer.

3.5 Development of flood events

The development of large events in both the pseudo-observations as well as the GWEX simulations (see 3.7.7) was studied by completing the hydrographs with precipitation data. Furthermore, the zero-degree line was included as indicator of the occurrence of snow fall or melt. In the 85-year pseudo-observation series, there is further the possibility to compare simulated runoff to measurements. The antecedent precipitation sum of large events was computed and plotted with hourly discharge, precipitation and the zero-degree line for both pseudo-observations and GWEX simulations.

The development of events was studied by selecting the largest discharge events and comparing them to the accumulated antecedent precipitation. Snow conditions and possible contributions of snow melt as indicated by the position of the zero-degree isotherm were tracked and related to the largest discharge events. In the following two large historic flood events (spring snowmelt flood 1999 and summer flood 2005) are presented for three example catchments to compare the different development over time in the same spatial setting.

Example catchment Dünnern: In the spring flood event 1999, the accumulated antecedent precipitation is around 100mm, the discharge peak of the event 70 m^3/s simulated and 210 m^3/s observed. During the month before the event there is snowmelt possible as indicated by the zero degree isotherm (Figure 32) and at the beginning of this period snow accumulation could have happened as indicated by the isotherm reaching into the range of the catchment area (horizontal dashed lines). In the summer flood event in 2005 the accumulated antecedent precipitation is around 200mm, the discharge peak of the event 14 m^3/s simulated and around 30 m^3/s observed. During the month before the event, no snowmelt is indicated by the zero degree isotherm (Figure 33). For the Dünnern catchment, the snow melt flood was larger than the summer flood even though the summer event had around double the amount of antecedent precipitation compared to the snow melt flood in 1999. Here, snow melt must have contributed considerably to the antecedent wetness and to discharge during the event.

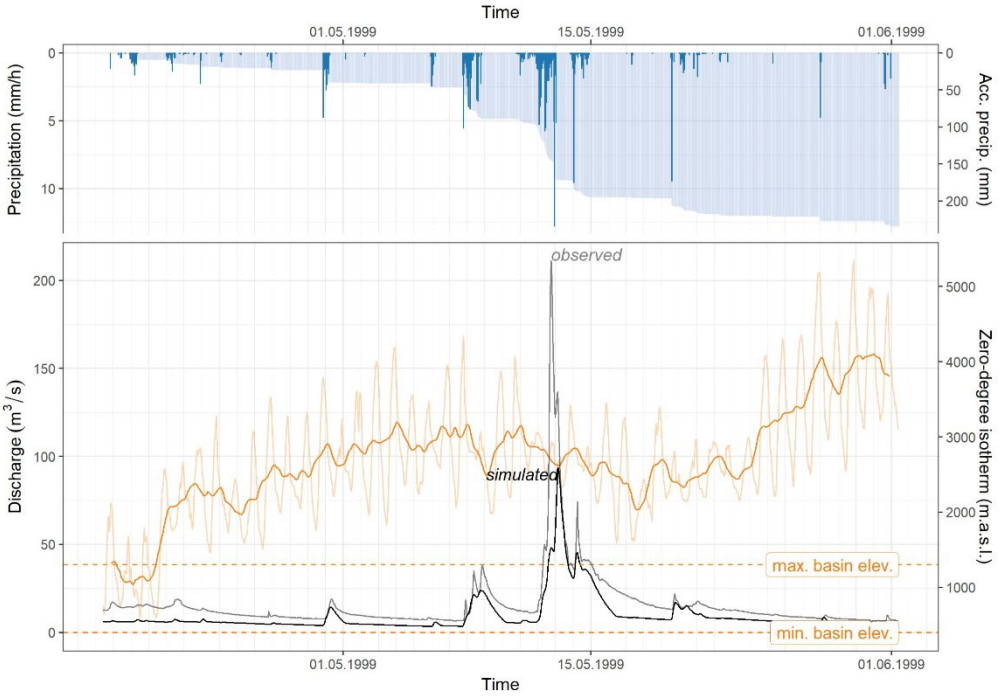


Figure 32 Historical spring flood 1999 at Dünnern. Hourly precipitation as well as accumulated precipitation in the upper panel, observed and simulated hydrographs as well as the zero degree isotherm (in light orange) and its 7-day moving average (dark orange) in the lower panel. For context on potential snow melt/accumulation, the catchment's elevation range is indicated with the horizontal dashed orange lines.

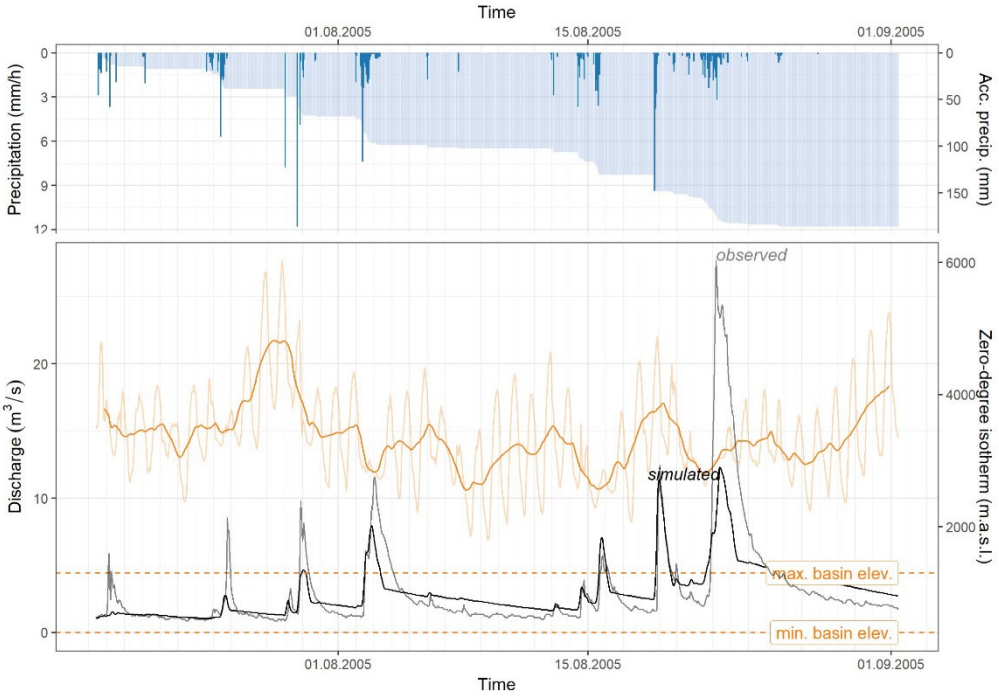


Figure 33 Historical summer flood 2005 at Dünnern. Hourly precipitation as well as accumulated precipitation in the upper panel, observed and simulated hydrographs as well as the zero degree isotherm (in light orange) and its 7-day moving average (dark orange) in the lower panel. For context on potential snow melt/accumulation, the catchment's elevation range is indicated with the horizontal dashed orange lines.

Example catchment Kander: In the spring flood 1999, the accumulated antecedent precipitation is around 100mm at the first peak and 150 mm at the second peak, the first simulated discharge peak of

the event is 140 m^3/s, the second 160 m^3/s. During the month before the event, the zero degree isotherm indicates that snow melt is possible (Figure 34).

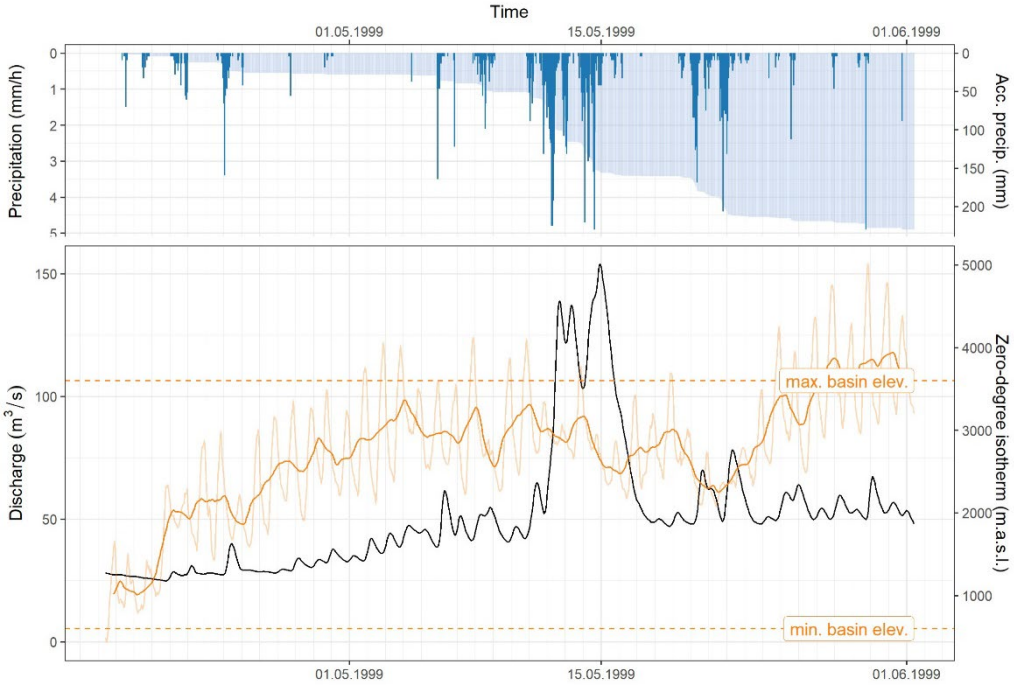


Figure 34 Historical event spring flood 1999 at Kander. Hourly precipitation as well as accumulated precipitation in the upper panel, observed and simulated hydrographs as well as the zero degree isotherm (in light orange) and its 7-day moving average (dark orange) in the lower panel. For context on potential snow melt/accumulation, the catchment's elevation range is indicated with the horizontal dashed orange lines.

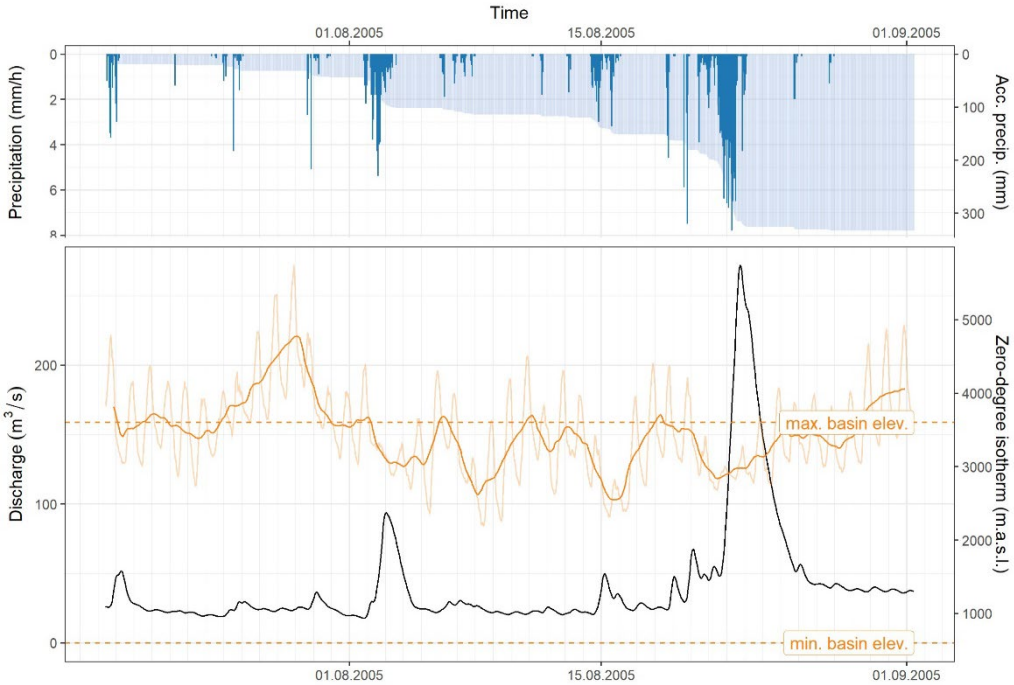


Figure 35 Historical event summer flood 20015 at Kander. Hourly precipitation as well as accumulated precipitation in the upper panel, observed and simulated hydrographs as well as the zero degree isotherm (in light orange) and its 7-day moving average (dark orange) in the lower panel. For context on potential snow melt/accumulation, the catchment's elevation range is indicated with the horizontal dashed orange lines.

At the beginning of this period, snow accumulation has likely occurred as indicated by the isotherm reaching into the catchment's elevation range. In the summer flood event in 2005, the accumulated antecedent precipitation is around 200 mm, the discharge peak of the event almost 300 m^3/s. During end of July, before the event, some snow accumulation and melt is likely, but not anymore during the days before the event in August. Different to the Dünnern catchment, for the Kander the spring melt event (1999) was smaller than the summer flood (2005), which developed from large amounts of precipitation in the preceding period, with potentially some addition of snowmelt pre-wetting the catchment (Figure 35).

Example catchment Kleine Emme: In the spring flood event 1999, the accumulated antecedent precipitation is around 120mm, the first simulated discharge peaks of the event are almost 300 m^3/s (both for simulated and observed discharge). During about two weeks before the event, is snowmelt likely (Figure 36Figure 34Figure 32), and at the beginning (first two weeks) of this period, snow accumulation might have occurred.

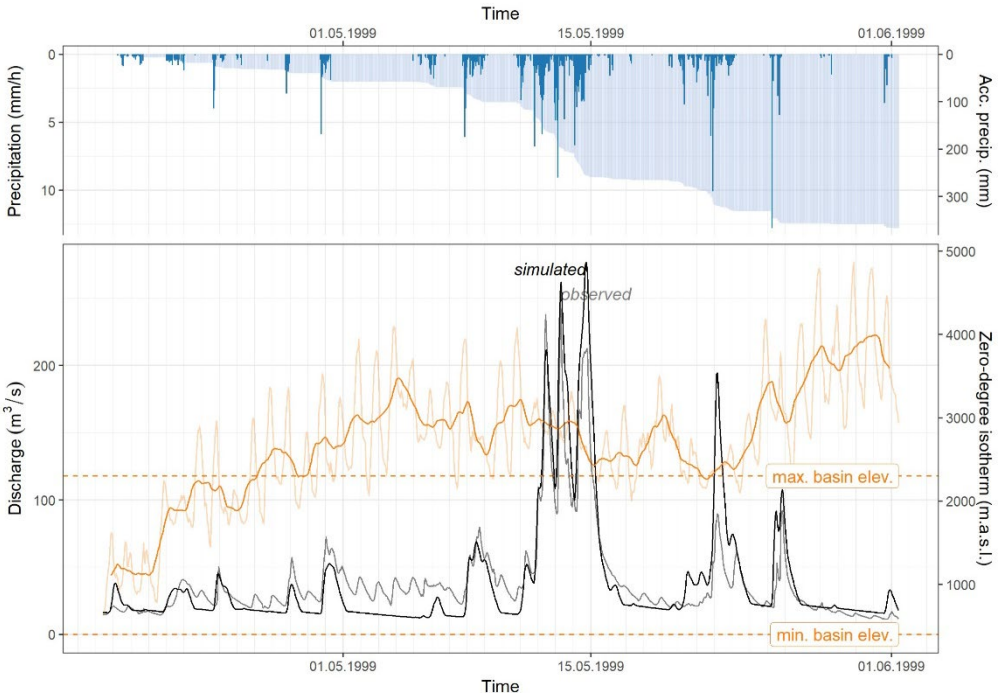


Figure 36 Historical spring flood 1999 at Kleine Emme. Hourly precipitation as well as accumulated precipitation in the upper panel, observed and simulated hydrographs as well as the zero degree isotherm (in light orange) and its 7-day moving average (dark orange) in the lower panel. For context on potential snow melt/accumulation, the catchment's elevation range is indicated with the horizontal dashed orange lines.

In the summer flood event in 2005, the cumulated antecedent precipitation is around 250 mm, the discharge peak of the event almost 650 m^3/s. There is no snow melt or snow accumulation during and before the event (Figure 37). For the Kleine Emme catchment, the spring melt event was about half the magnitude of the summer flood. The latter developed only from large amounts of precipitation in the preceding period.

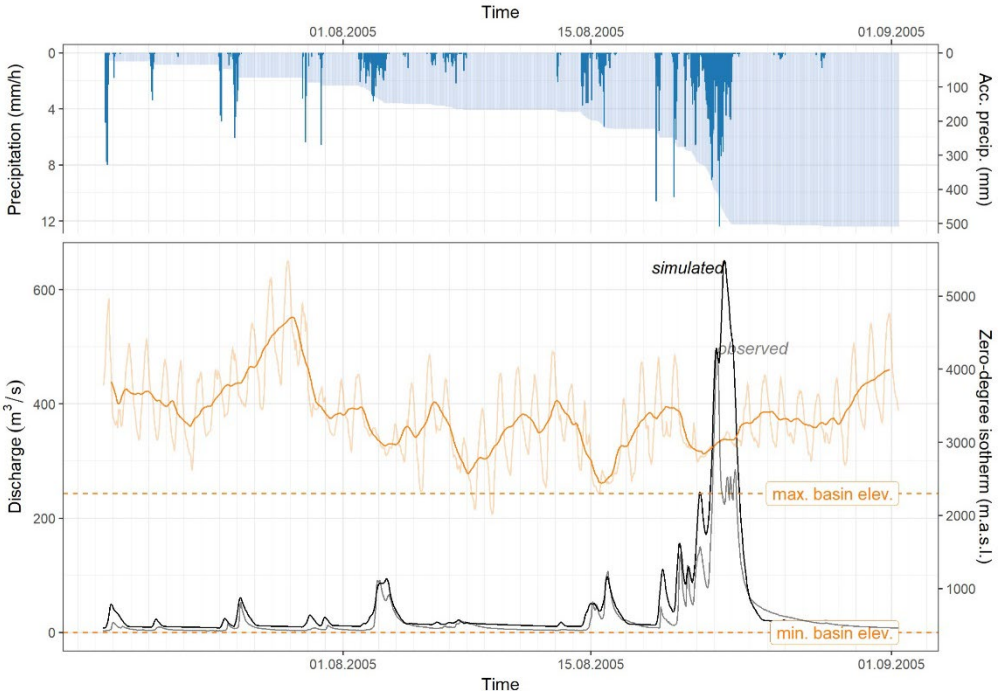


Figure 37 Historical summer flood 2005 at Kleine Emme. Hourly precipitation as well as accumulated precipitation in the upper panel, observed and simulated hydrographs as well as the zero degree isotherm (in light orange) and its 7-day moving average (dark orange) in the lower panel. For context on potential snow melt/accumulation, the catchment's elevation range is indicated with the horizontal dashed orange lines.

Example catchment Lütschine: In the spring flood 1999, the accumulated antecedent precipitation is around 50 mm, the first simulated discharge peaks of the event were about 140 m^3/s (both for simulated and observed discharge). During two weeks before the event there is snowmelt possible as indicated by the zero degree isotherm (Figure 38Figure 32).

In the summer flood event in 2005 the accumulated antecedent precipitation is around 200 mm, the discharge peak of the event almost 300 m^3/s. During and before the event there is no particular snow accumulation and melt evident as indicated by the zero degree isotherm (Figure 39). For the Lütschine catchment the spring melt event was about half of the summer flood, which developed only from large amounts of precipitation in the preceding period.

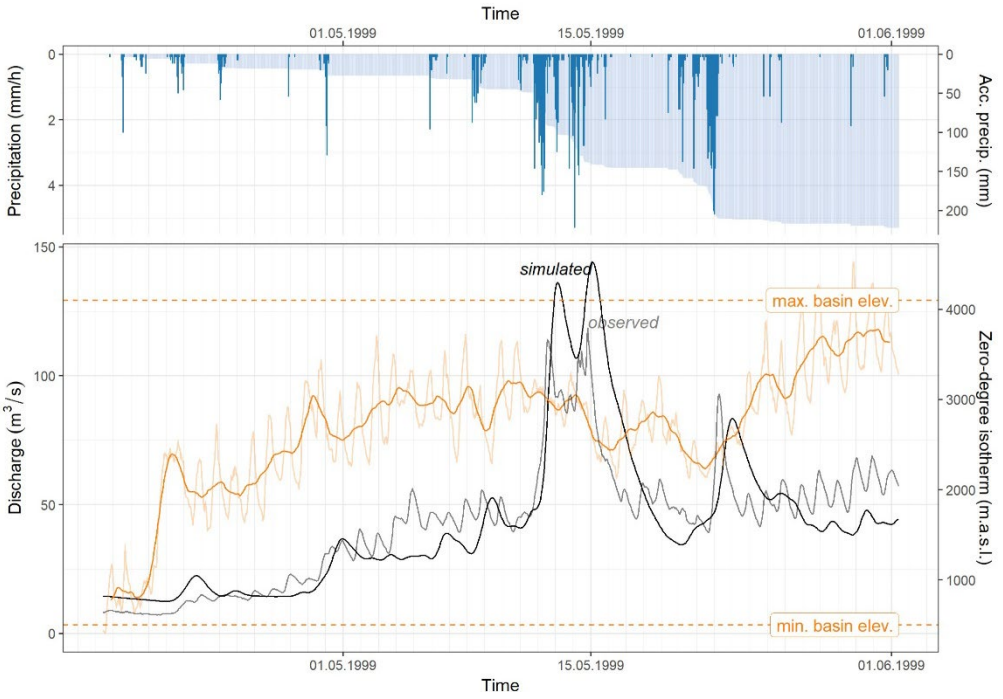


Figure 38 Historical spring snow melt flood 1999 at Lütschine. Hourly precipitation as well as accumulated precipitation in the upper panel, observed and simulated hydrographs as well as the zero degree isotherm (in light orange) and its 7-day moving average (dark orange) in the lower panel. For context on potential snow melt/accumulation, the catchment's elevation range is indicated with the horizontal dashed orange lines.

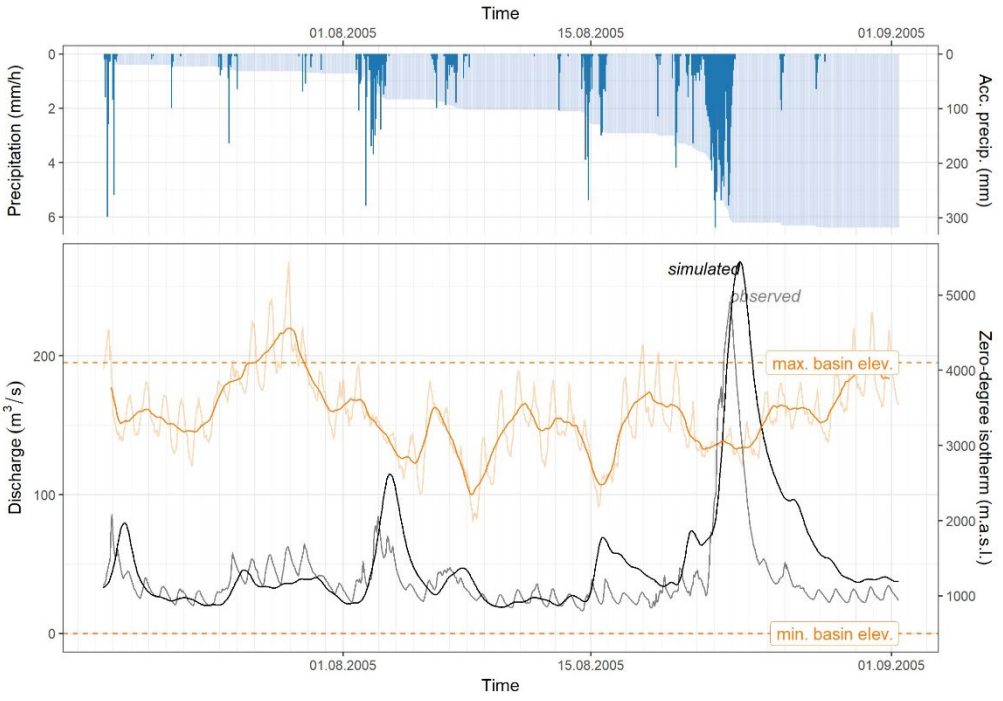


Figure 39 Historical event summer flood 2005 at Lütschine. The Upper panel shows hourly precipitation as well as accumulated precipitation, the lower panel shows the observed and simulated hydrographs as well as the zero degree isotherm (in light orange) and its 7-day moving average (dark orange). For context on potential snow melt/accumulation, the catchment's elevation range is indicated with the horizontal dashed orange lines.

Hence, measurements and simulations compare well for the events studied, and the temporal development of the simulated variables is reasonable. The behaviour of the zero degree isotherm

indicated that the processes leading to snow melt floods are captured in the simulations and accumulated precipitation is considerably higher for summer floods than for snow melt floods.

3.6 Disaggregated weather input: spatial pattern and validation with observations

The quality and plausibility of the disaggregated weather inputs (pseudo-observations 1930–2014) were assessed by looking at 1) spatial patterns of precipitation events compared to historically observed events, 2) the performance of hydrological model simulations, i.e. verifying whether HBV was able to translate the given precipitation input into plausible runoff simulations.

3.6.1 Spatial pattern of disaggregated weather inputs and runoff response

The spatial pattern of observed historical flood events were compared to both the precipitation patterns and the resulting runoff patterns. For this three major historical runoff events were selected: May 1999, August 2005 and August 2007. The event of 1999 was a typical snow melt event, while the events of 2005 and 2007 were flood events caused by long and intense precipitation. The resulting patterns are described in more detail in the following, the plausibility of these patterns was confirmed by MeteoSwiss (personal communication with Sophie Fukutome and Andreas Fischer).

Event 05-1999: Most precipitation in the 1999 event fell in the area around Lake Zurich and the accumulated precipitation increases for all catchments when summing over a larger aggregation window (Figure 40 and Figure 41). However, only a few catchments respond with high specific discharge to large precipitation (Figure 42). Specific discharge is rather relatively homogenous in space and more guided by the mean elevation of the catchments. Snow melt occurred mainly in the transient zones, where there was still snow, but not in the higher alpine catchments. The specific discharge at the TPs is only little influenced by the catchments with the highest specific discharges.

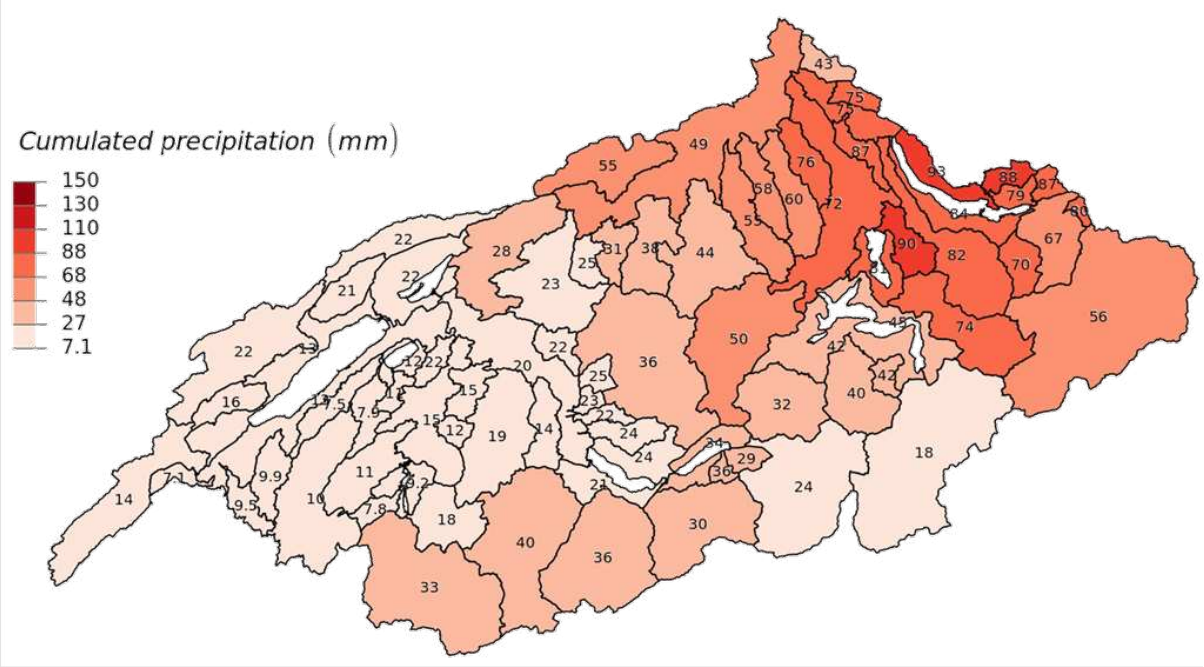


Figure 40 Antecedent accumulated precipitation 24 hours prior to the runoff event of 05-1999. The numbers in the catchment shapes indicate the exact precipitation sums (mm) over 24h.

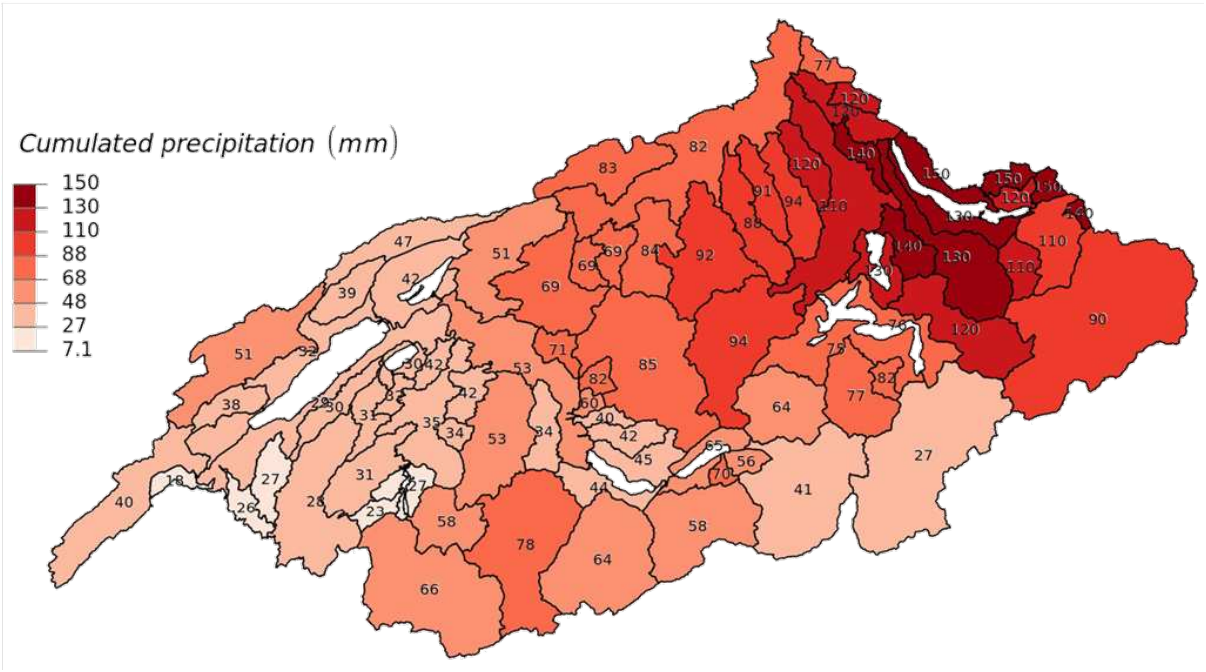


Figure 41 Antecedent accumulated precipitation 72 hours prior to the runoff event of 05-1999. The numbers in the catchment shapes indicate the exact precipitation sums (mm) over 72h.

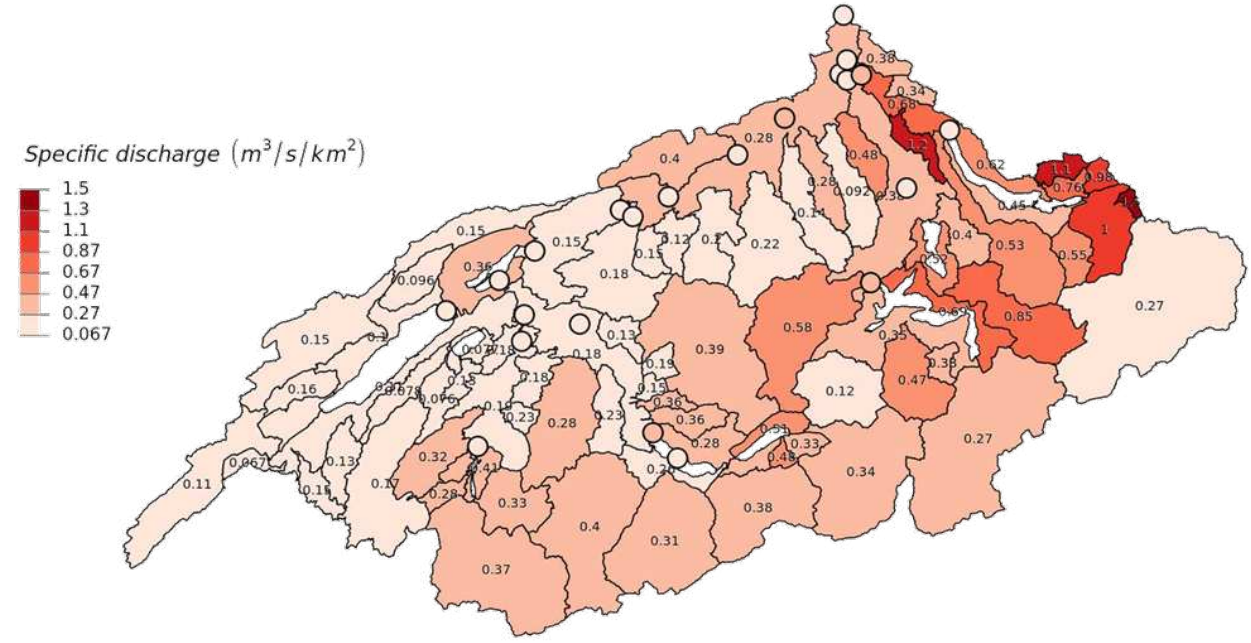


Figure 42 Specific peak discharge for the event of 05-1999 for all catchments and TP points (dots in map).

Event 08-2005: The accumulated precipitation for 72 h reached values of more than 200mm for some catchments. Generally, this event is the largest of the three selected events. The focus area of the summer event in 2005 was in the alpine region and in the central part of Switzerland, when looking at the precipitation patterns for the cumulated precipitation 24 h and 72h (Figure 43 and Figure 44). For this event, the runoff response followed the pattern of the antecedent precipitation.

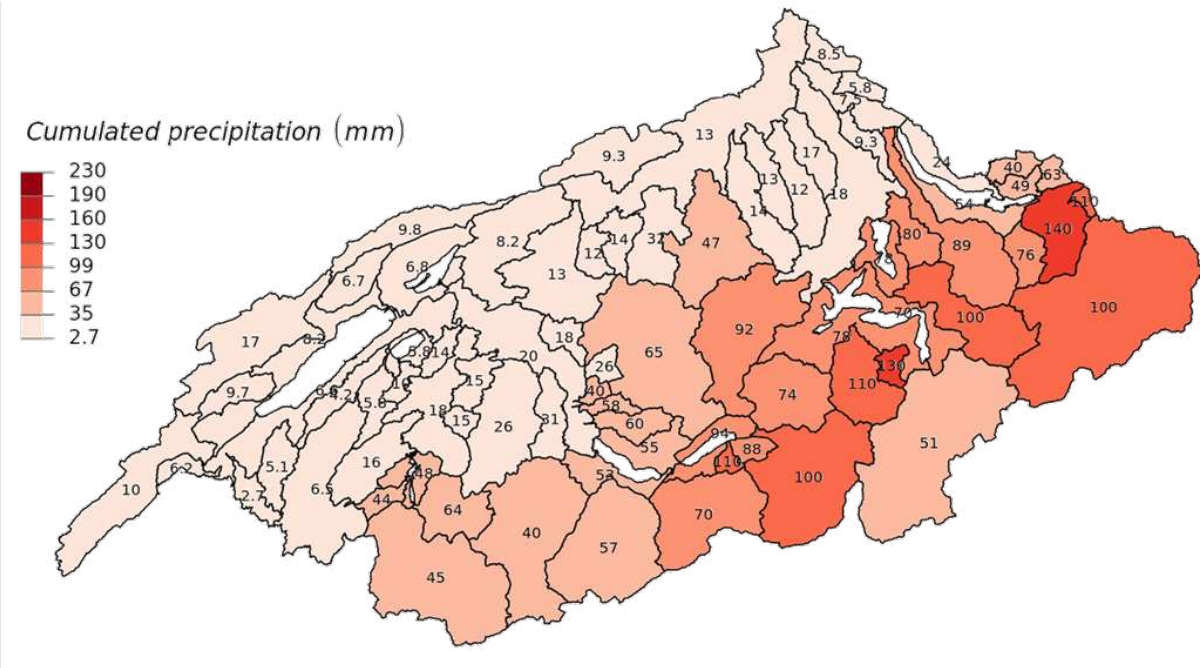


Figure 43 Antecedent accumulated precipitation 24 h prior to the runoff event of 08-2005. The numbers in the catchment shapes indicate the exact precipitation sum (mm) over 24 h.

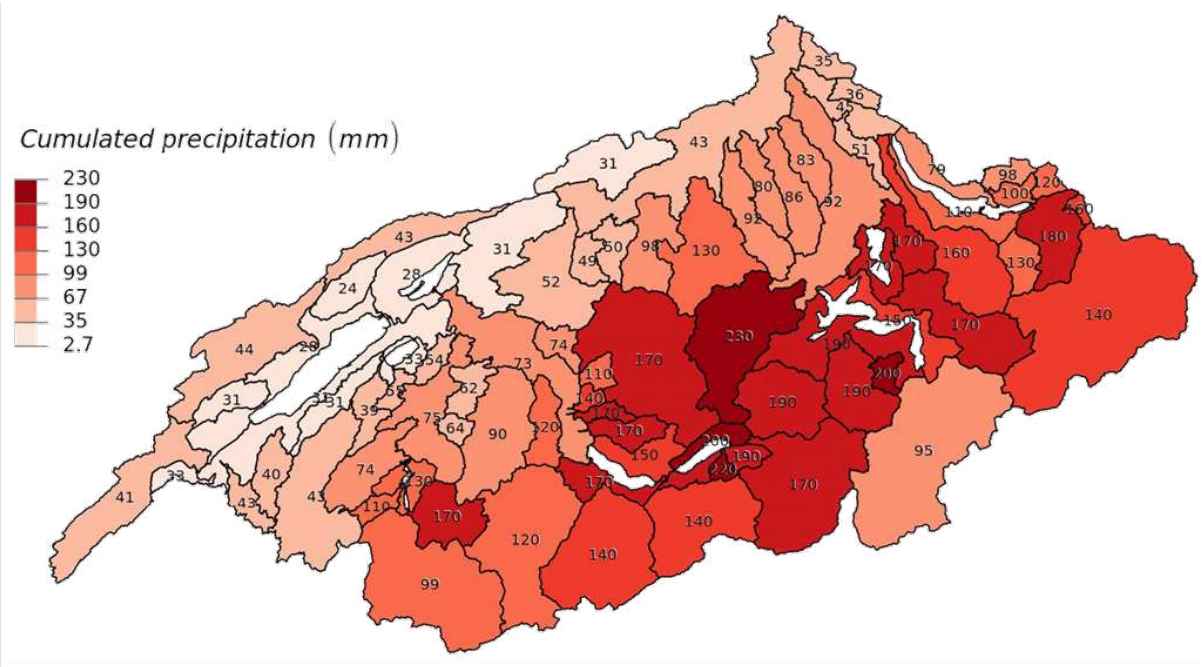


Figure 44 Antecedent accumulated precipitation 72 h prior to the runoff event of 08-2005. The numbers in the catchment shapes indicate the exact precipitation sums (mm) over 72 h.

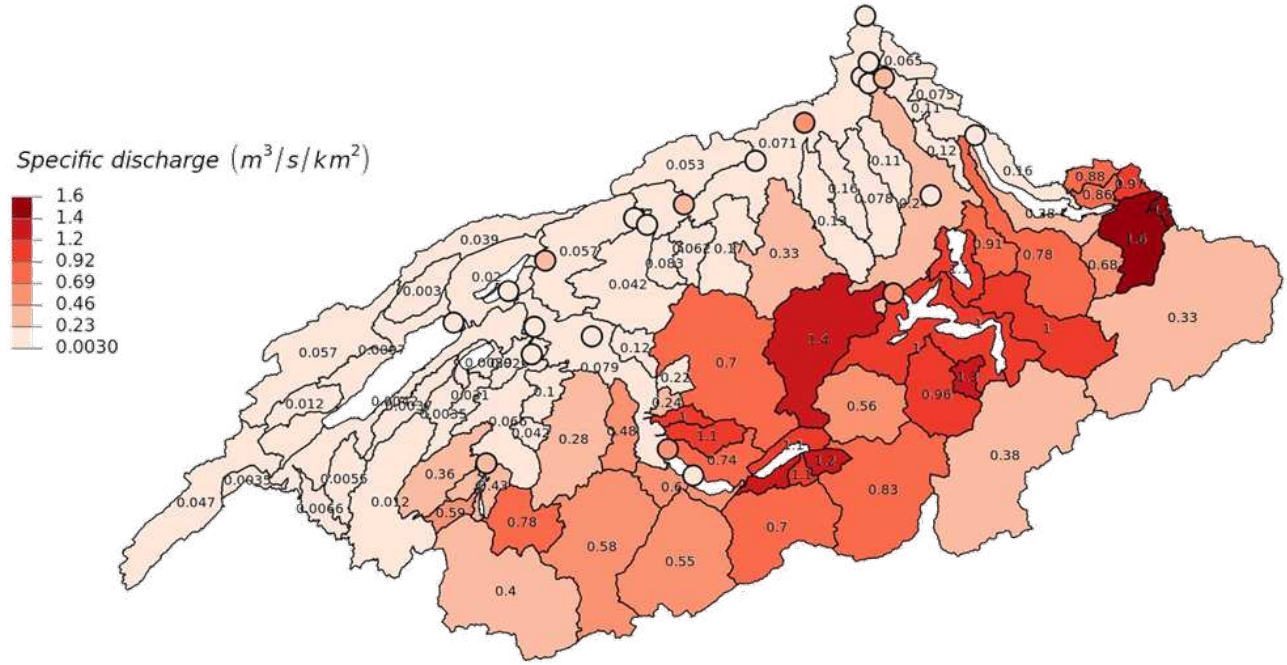


Figure 45 Specific peak discharge for the event of 08-2005 for all catchments and TP points (dots in map).

Event 08-2007: The focus of the event is more in the western and north eastern part of the Aare River basin, while the alpine parts received less precipitation for the aggregation period of 24 hours. Instead for 72 hours, the full Aare basin receives almost uniformly precipitation with slightly lower values in the high alpine catchments (Linth, Aare-Brienz and Reuss). The specific discharge at the TPs is only little influenced by the catchments with the highest specific peak discharges, which can be explained with the regulation of the lakes in the upper catchments.

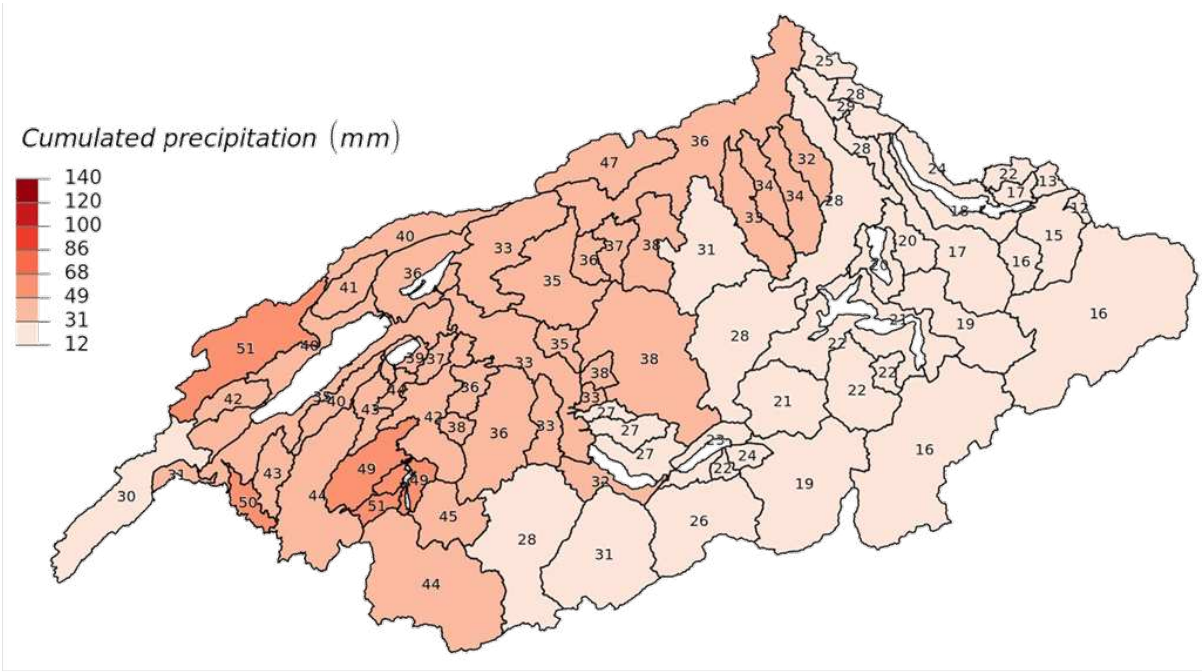


Figure 46 Antecedent accumulated precipitation 24 h prior to the runoff event of 08-2007. The numbers in the catchment shapes indicate the exact precipitation sum (mm) over 24 h.

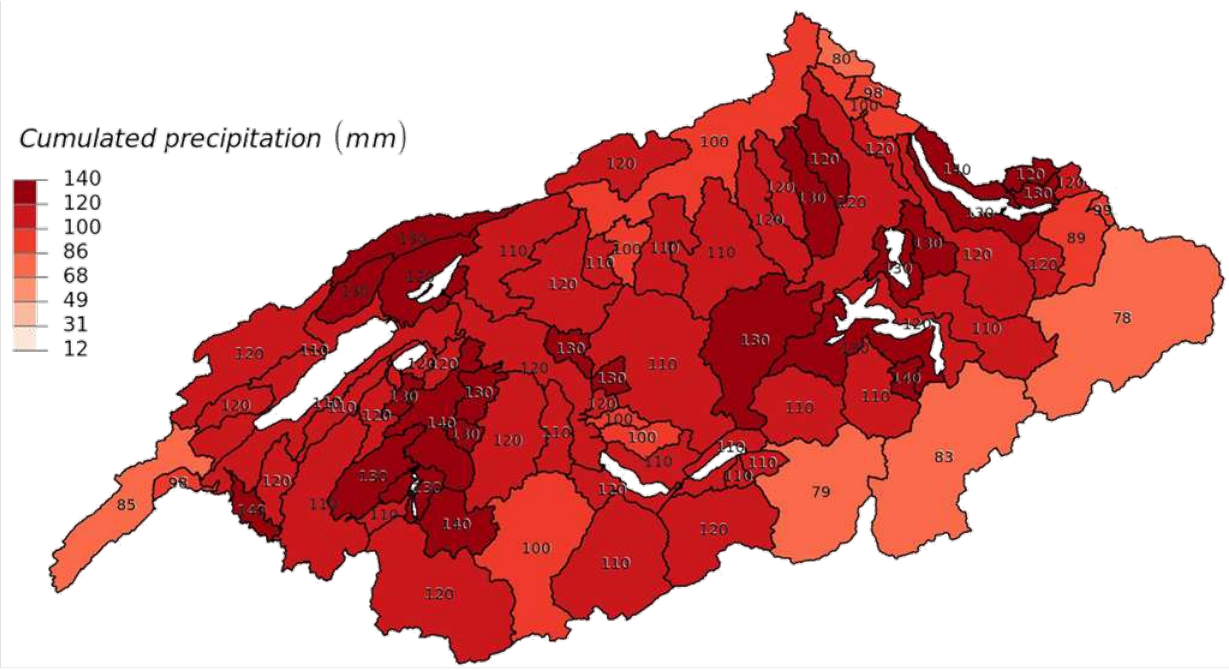


Figure 47 Antecedent accumulated precipitation 72 h prior to the runoff event of 08-2007. The numbers in the catchment shapes indicate the exact precipitation sums (mm) over 72 h.

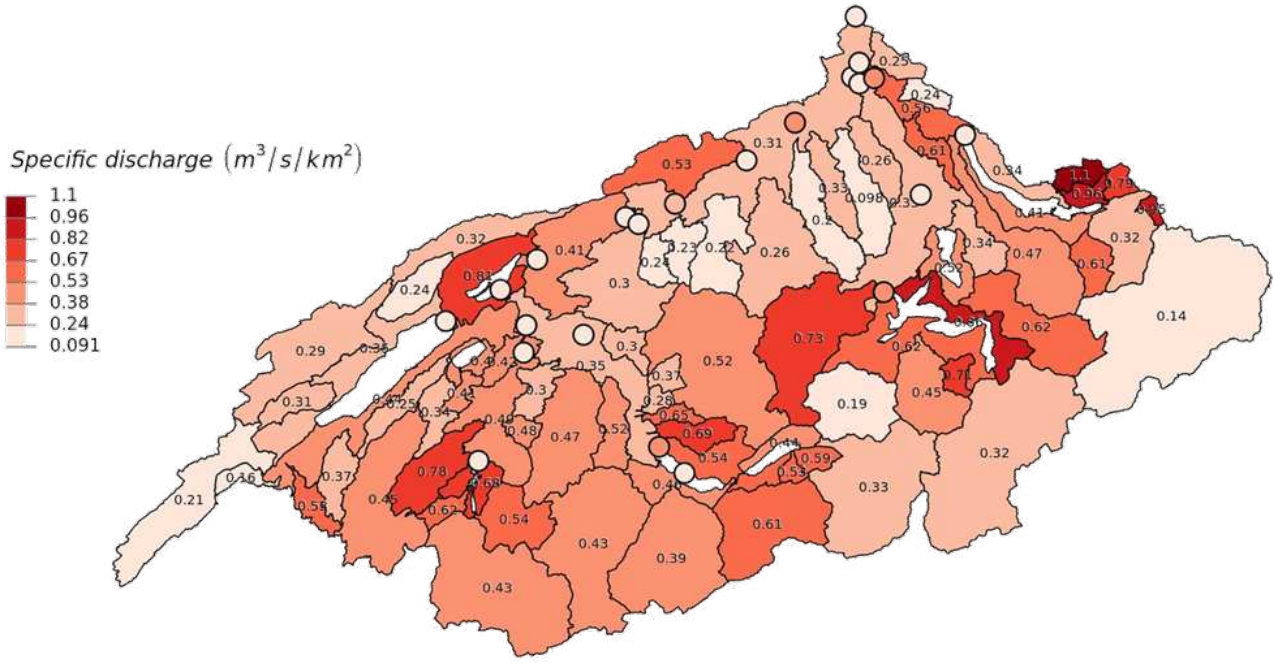


Figure 48 Specific peak discharge for the event of 08-2007 for all catchments and TPs (dots in map).

3.6.2 Performance of hydrological model simulations

Performance of both the HBV (individual catchments) and RS Minerve (total system of the Aare River basin) models was tested using the 1930–2014 pseudo-observations as input data. The resulting runoff simulations were then compared to runoff observations at hourly and daily resolution, which are available for 44 HBV catchments and 10 RS Minerve nodes.

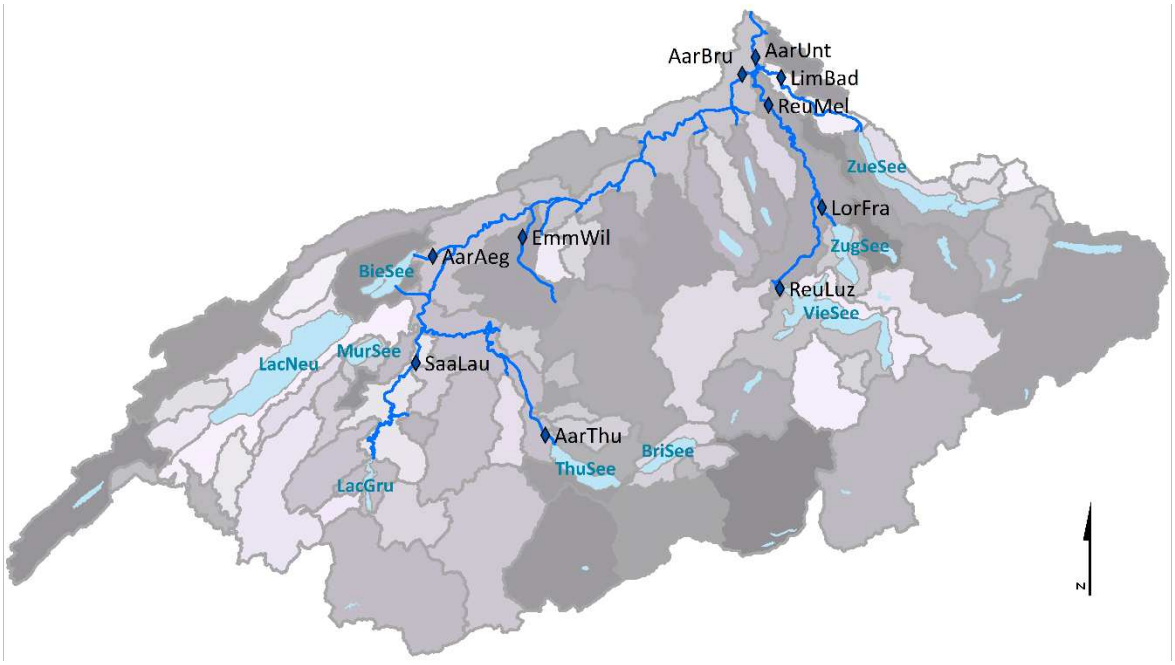


Figure 49 Location of discharge gauges used for validation of discharge simulations for the total system (blue diamonds); in light blue are indicated the 9 simulated lakes (see section 3.7.6).

The performance of the hydrological model was evaluated both for the entire observation period as well as for individual events. Flow duration curves (FDCs) from the real observations and the pseudo-observations as well as efficiency criteria (Table 3). FDC were calculated to examine whether the models can sufficiently well represent the occurrence of flows of different magnitude. The efficiency criteria evaluate the simulations from a slightly different perspective, each criterion putting different emphasis on certain parts of the hydrograph.

Model performance was also evaluated based on the annual maximum floods (AMFs), i.e. event based. Here, no efficiency scores were applied but simulated peak runoff was compared to observed peak runoff, once in absolute terms (mm/h) and once in relative terms ($AMF_{i,sim} – AMF_{i,obs}/AMF_{i,obs}$), where the latter allows better for comparing the catchments to each other.

Table 3 Performance evaluation for the entire observation period using different efficiency criteria. All efficiency criteria shown have a range of]-∞, 1], with a perfect match at 1.

Efficiency criterion	Equation	Interpretation
Nash-Sutcliffe (NSE) Nash & Sutcliffe, 1970	$\mathbf{1} - \sum \frac{(\boldsymbol{Q_{obs}} - \boldsymbol{Q_{sim}})^2}{(\boldsymbol{Q_{obs}} - \bar{\boldsymbol{Q}}_{\boldsymbol{obs}})^2}$	Evaluates water balance and dynamic of the hydrograph with emphasis on the high flows.
Kling-Gupta (KGE) Gupta et al., 2009	$\beta = \frac{\mu_{sim}}{\mu_{obs}}$ $\alpha = \frac{\sigma_{sim}}{\sigma_{obs}}$ $r_p = \frac{\sum_{i=1}^{n}(Q_{obs}(i) - \mu_{obs})(Q_{sim}(i) - \mu_{sim})}{\sum_{i=1}^{n}\sqrt{(Q_{obs}(i) - \mu_{obs})^2\,(Q_{sim}(i) - \mu_{sim})^2}}$ $\mathbf{1} - \sqrt{(\boldsymbol{\beta} - \mathbf{1})^2 + (\boldsymbol{\alpha_{KG}} - \mathbf{1})^2 + (\boldsymbol{r_p} - \mathbf{1})^2}$	Evaluates water balance and dynamic of the hydrograph considering the entire hydrograph.
Modified Kling-Gupta (KGE_MOD) Pool et al., 2018	$\alpha_{NP} = 1 - \frac{1}{2}\sum_{k=1}^{n}\left\lvert\frac{Q_{sim}(I(k))}{n\bar{Q}_{sim}} - \frac{Q_{obs}(J(k))}{n\bar{Q}_{obs}}\right\rvert$ $r_s = \frac{\sum_{i=1}^{n}(R_{obs}(i) - \bar{R}_{obs})(R_{sim} - \bar{R}_{sim})}{\sqrt{(\sum_{i=1}^{n}(R_{obs}(i) - \bar{R}_{obs})^2) - (\sum_{i=1}^{n}(R_{sim}(i) - \bar{R}_{sim})^2)}}$ $\mathbf{1} - \sqrt{(\boldsymbol{\beta} - \mathbf{1})^2 + (\boldsymbol{\alpha_{NP}} - \mathbf{1})^2 + (\boldsymbol{r_s} - \mathbf{1})^2}$	Evaluates water balance and dynamic of the hydrograph considering the entire hydrograph in a non-parametric manner.

Flow duration curves

Looking at the entire observation period, the FDCs derived from the simulations were very similar to the ones from the observations for most catchments (Figure 50). Larger differences were found for the catchments Simme (SimLat) and Chise (ChiFre), where for the Simme catchment this could be explained by the observations that are not entirely credible (discharge needs to be summed up from two stations) while the difference at the Chise catchment can be explained with karst. The differences between the three parameterizations were very small.

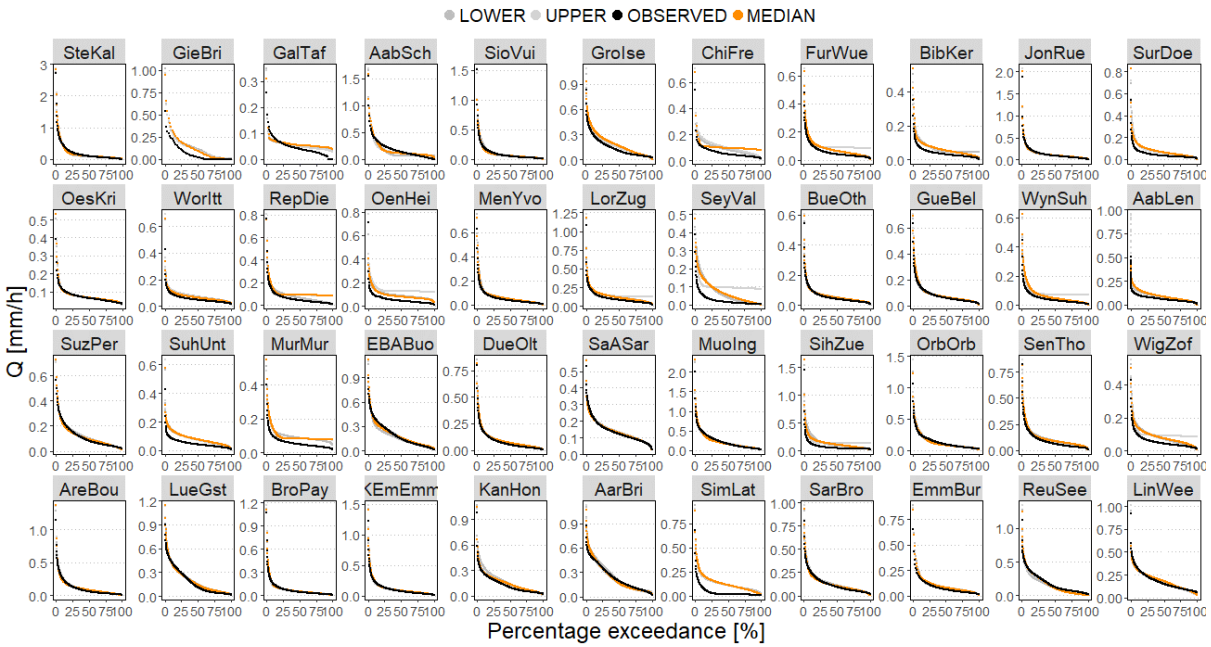


Figure 50 FDCs for the simulations based on pseudo-observations compared to real observations (hourly data) for all parameterizations (LOWER, MEDIAN, UPPER) of the hydrological model. The catchments are sorted by catchment size starting with the smallest in the upper left corner and ending with the largest in the lower right corner of the plot.

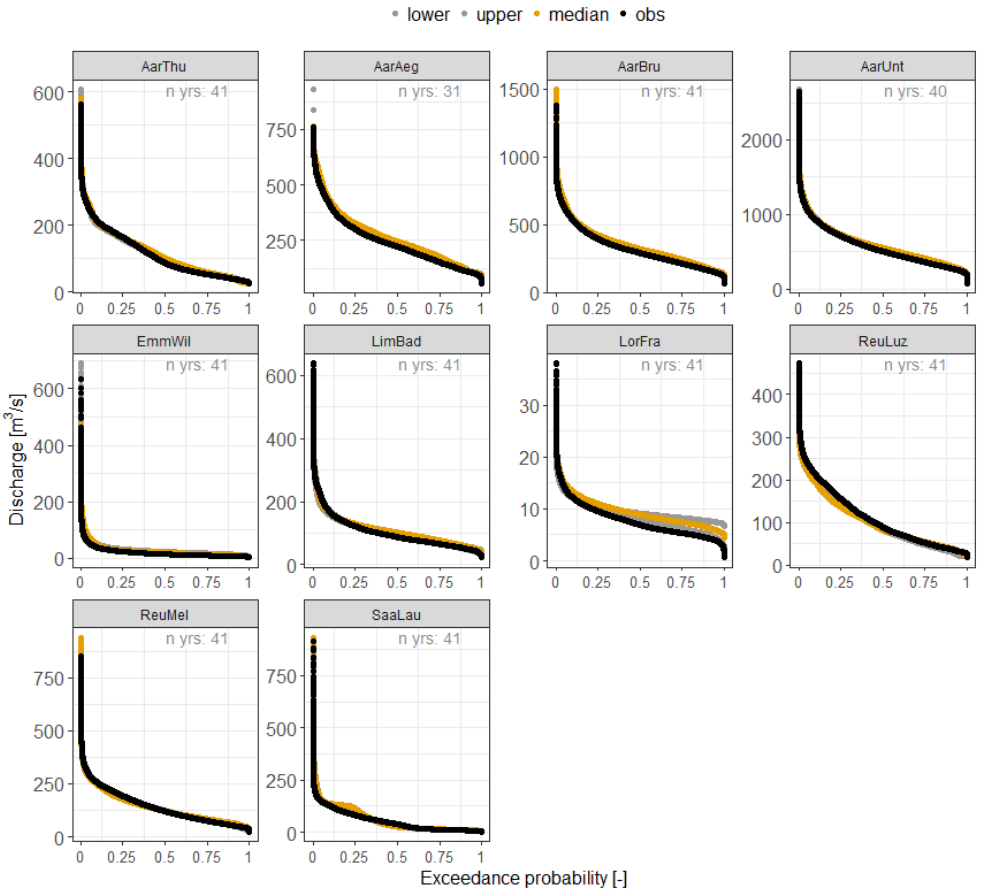


Figure 51 FDCs for simulated discharge based on pseudo-observations compared to real observations (hourly data) for all parameterizations (lower, median, and upper) of the hydrological model at the ten gauged points of the total system.

FDCs curves of hourly discharge for the 10 nodes used for validating the total system, i.e. for about 30 to 40 years depending on the corresponding gauging station, were derived from the simulations and were found to be very similar to the ones derived from observations (***Figure 51***). Larger differences were found for the Lorze (LorReu), with a systematic overestimation of more frequent flows, and an underestimation of the high flows. For LorReu the best fit was found applying the upper parameter set, and the biggest differences between the three parameterizations are found, while generally these are very small for all other analyzed points.

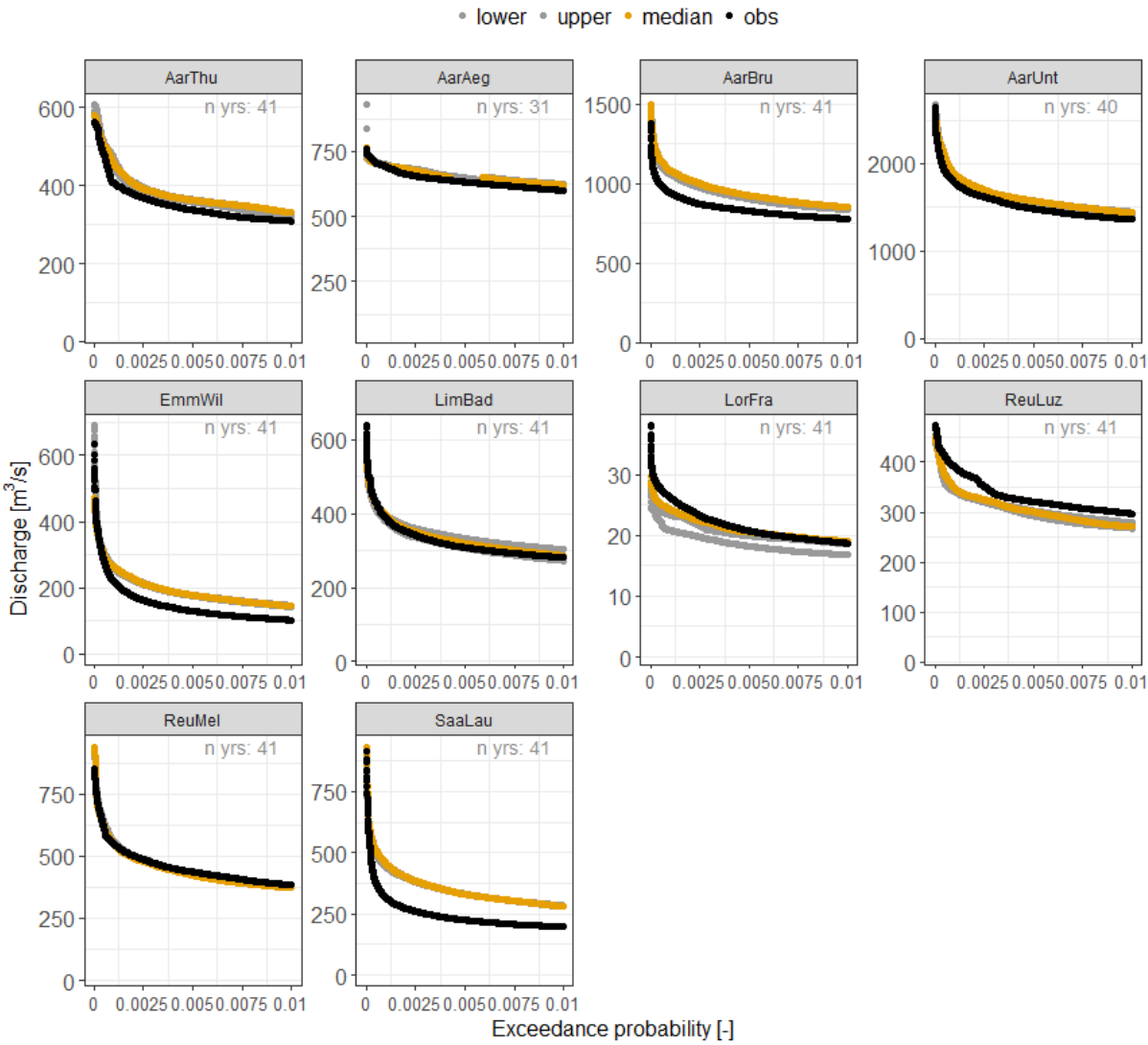


Figure 52 FDCs of the top 10% of simulated flows based on pseudo-observations compared to real observations (hourly data) for all parameterizations (lower, median, and upper) of the hydrological model at the ten gauged points of the total system. Model performance for entire observation period.

Plotting FDCs curves of the top 10% of the simulated flows (Figure 52) allowed to see a general tendency to overestimate discharge for the Aare up to Aegerten (AarAeg) and again downstream for the Aare up to and including Brugg (AarBru), reflecting the propagation of the overestimation found for the Saane (SaaLau) and the Emme (EmmWil) respectively.

All efficiency criteria reached acceptable ranges for the catchments examined (Figure 53). For some catchments the simulations were poor, however, i.e. lower than 0. A closer look revealed that many

of them are karstic and some contain regulated lakes. The HBV model was not set up explicitly for any of these two cases, hence here the poor model performance can be explained by HBV's model structure rather than by the disaggregated weather inputs. The different parameterizations (lower, median, upper) had no major effect on model performance apart from the upper parameterization, which showed a tendency to a larger spread than the other two. However, there was no difference in the median of the distributions (boxes in plot) for the different parameterizations (Figure 53).

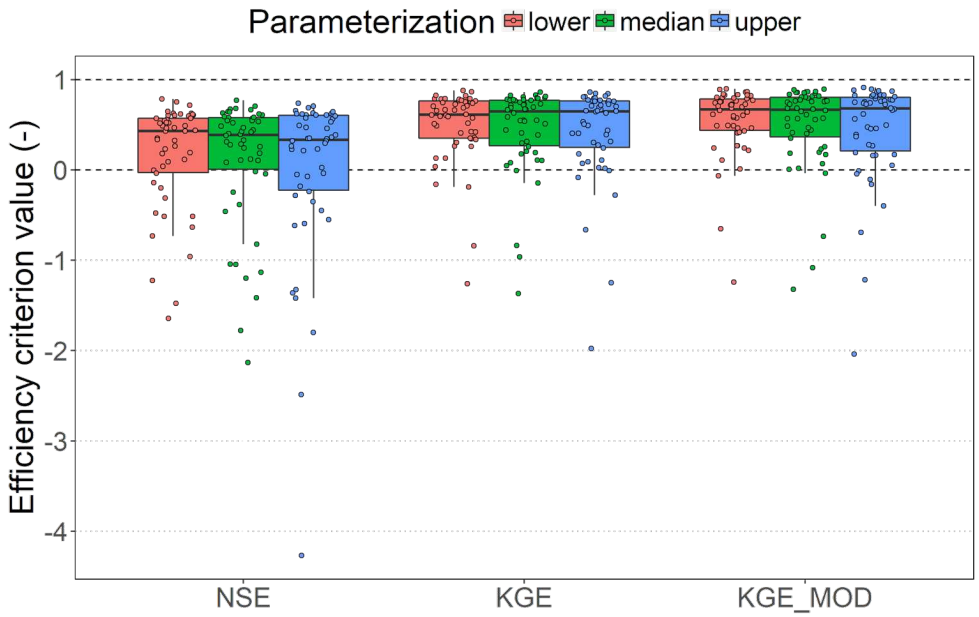


Figure 53 Efficiency criteria for all catchments for the median, lower and upper parameterizations. The dots show the efficiency criteria for each catchment. The dashed line at zero indicates the efficiency measure's specific benchmark, for NSE for instance using mean flow as predictor.

Efficiency criteria total system

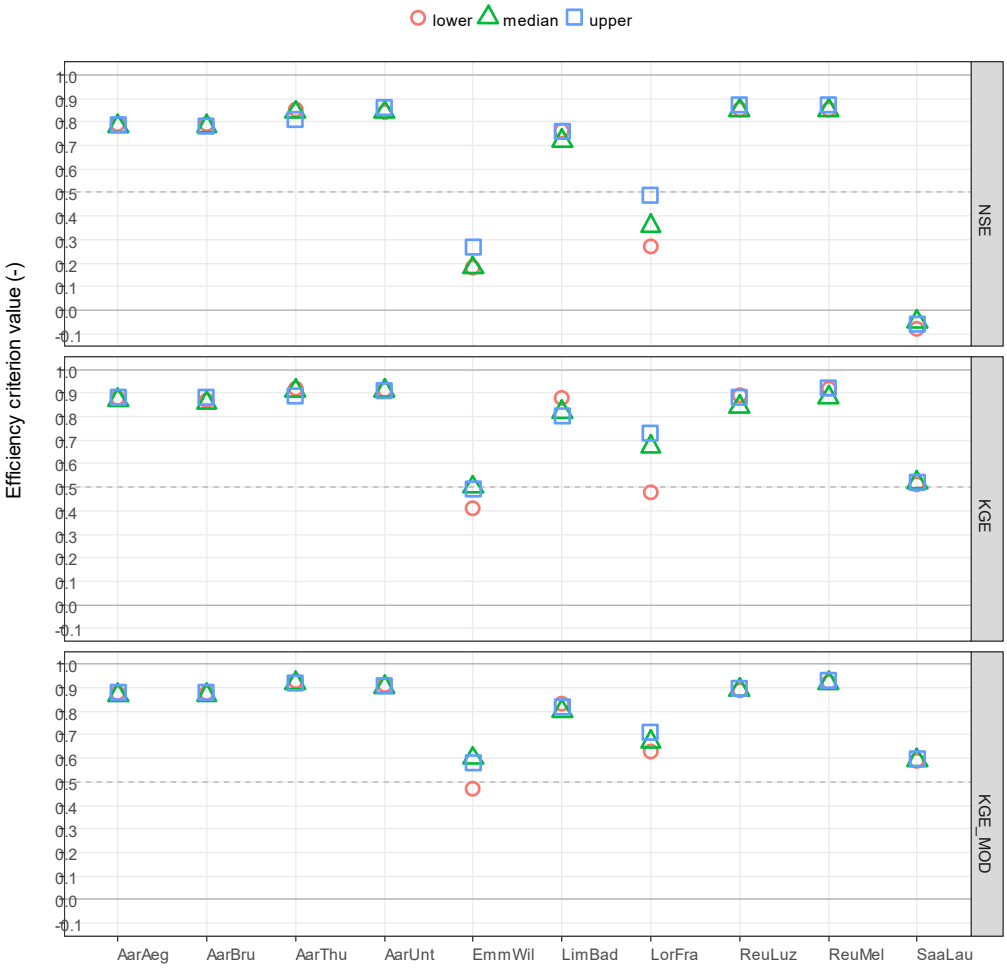


Figure 54 Model efficiency for simulated hourly discharge. Top to bottom: Nash-Sutcliffe-Efficiency (NSE), Kling-Gupta-Efficiency (KGE), and a non-parametric version of the Kling-Gupta-Efficiency (KGE_MOD). Different symbols indicate the HBV parameterizations (lower, median, upper).

For the period in which discharge observations are available at high temporal resolution (roughly 1974–2014), efficiency criteria were calculated using hourly data. Resulting efficiencies showed good to very good agreement between observations and simulations (efficiency criteria between 0.72 and 0.93) for all gauges at the Aare, Reuss and Limmat Rivers (Figure 54).

Three gauges show poorer performance, namely these at the Emme, Lorze, and Saane Rivers. For validation of the entire period (1930–2014) simulated discharge was aggregated to daily mean values. Model performance is similar or improved using these data (Figure 55).

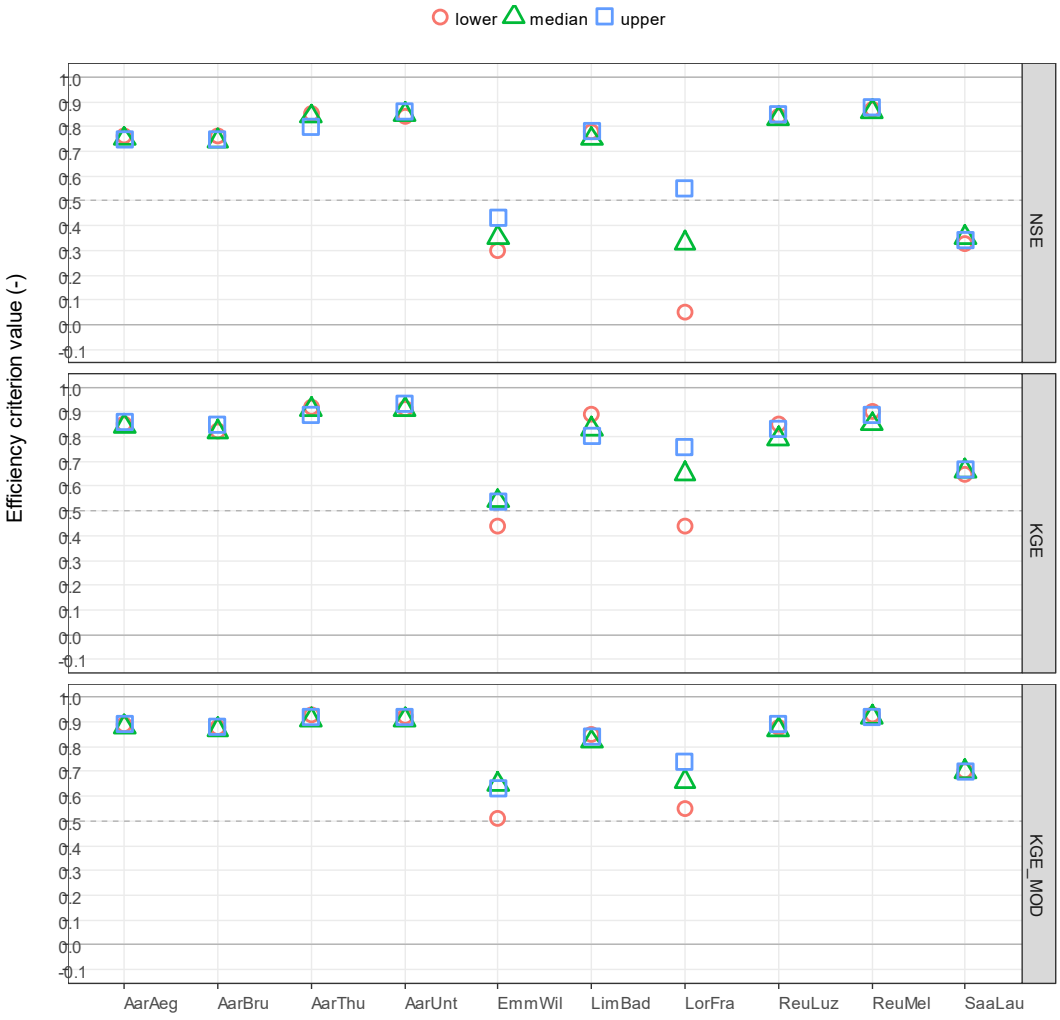


Figure 55 Model efficiency for simulated daily discharge. Top to bottom: Nash-Sutcliffe-Efficiency (NSE), Kling-Gupta-Efficiency (KGE), and a non-parametric version of the Kling-Gupta-Efficiency (KGE_MOD). Different symbols indicate the HBV parameter set.

Event based evaluation of model performance

The event based analysis the absolute differences of most of peaks were in a range +/- 1mm (Figure 57) for both the largest peak and the largest volume events. For some of the catchments there were no observations available for all events, which is visible in the gaps in the plots. The differences were smaller for the larger catchments and greater for the smaller catchments. The largest difference could be found for the Steinenbach (SteKal) catchment ([-3mm, +2mm]), which is the smallest catchment considered in the Aare the system and hence contributes only little. Hence, for the full system the absolute errors are in an acceptable range and the simulations based on the pseudo-observations were close to the real observations. Looking at the different events there was no systematic pattern apparent that one event was particularly worse of better than the other events simulated, which means that the model performed equally well for each event.

The events that were analyzed are shown for all catchments in Figure 56. Here we expect the median parameterization of the HBV model to be the one that represents the observation closest, hence the analysis of peak volumes is focused on the comparison between the simulations of the median parameterized HBV model and the observations.

The response of the individual catchments varies for the same event, both in time of occurrence as well as amount. Events for which there are no data were not included in the summary statistics over all catchments. For the events the peak volumes were compared in absolute amounts (mm) and relative to the peak volume of the observation ($(Q_{peak\ sim} - Q_{peak\ obs})/Q_{peak\ obs}$) (-).

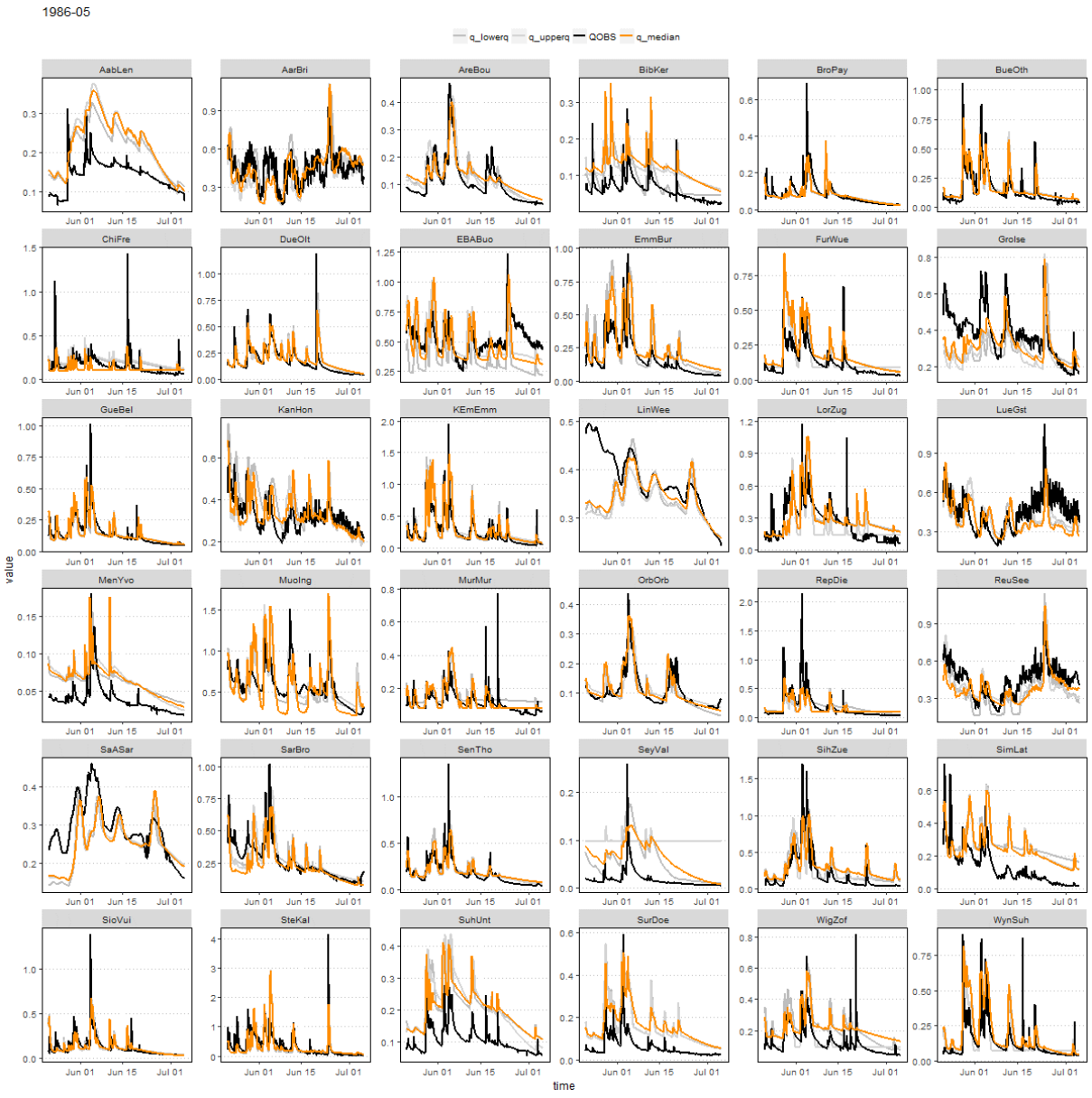


Figure 56 Hydrographs of one of the evaluated events (1986-05). The black line is observed discharge, the orange line simulation using the median parameter set, the grey lines simulations using the upper and lower parameter sets, respectively.

The relative volume errors between simulations with median parameterization are shown for both the largest volume events and the largest peak events (Figure 57). For both kinds of events, the error is decreasing with increasing catchment size. In the highest peak events, the simulation error of the Chise (ChiFre) catchment appears very large, but the box for this catchment is only made from two points (two events) and hence the large span of the box is misleading. For the largest volume events, the error is even smaller for all catchments than the error for the largest peak events, however the focus of the

evaluation was on the difference of discharge volumes at the flow peak rather than on the volume of the full event.

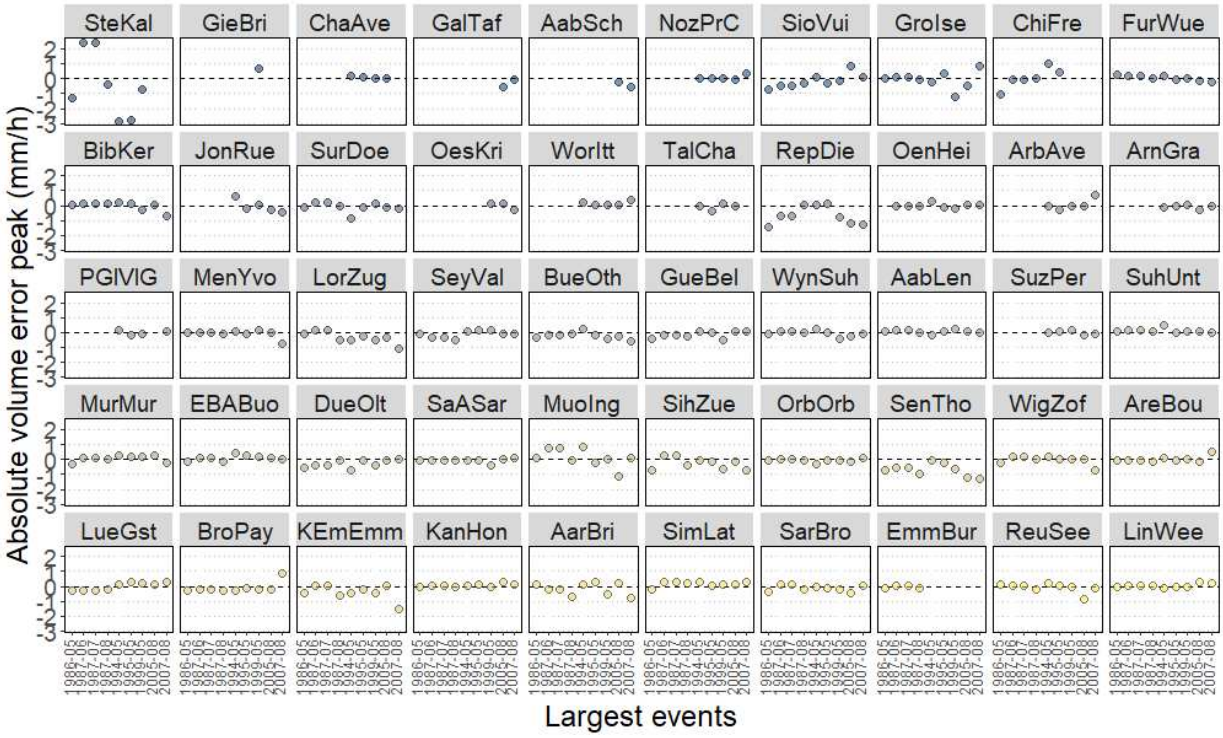


Figure 57 Absolute volume errors of the peak (mm/h) for each catchment and for each event (largest peak and volume events). The catchments are sorted by catchment size and the colors match with the map in Figure 58.

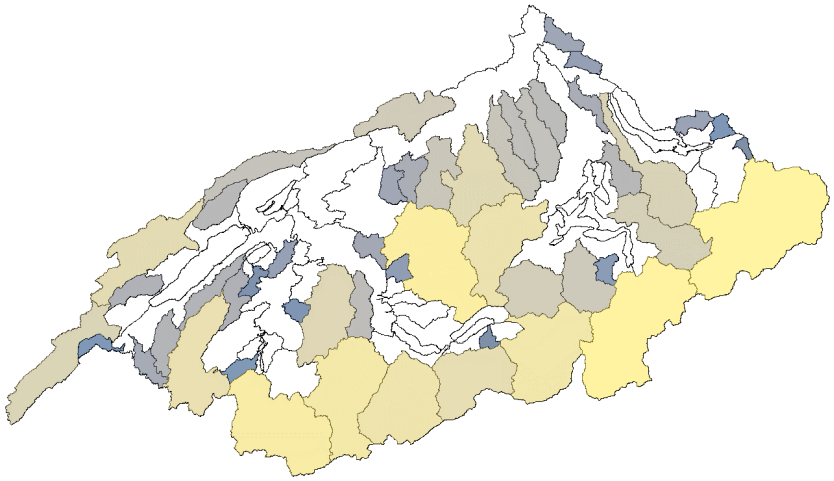


Figure 58 Catchments available to compare the results from pseudo-observation-based to simulations based on real observations. The color indicates the catchment size, with bluer hues for smaller and more yellow hues for larger catchments.

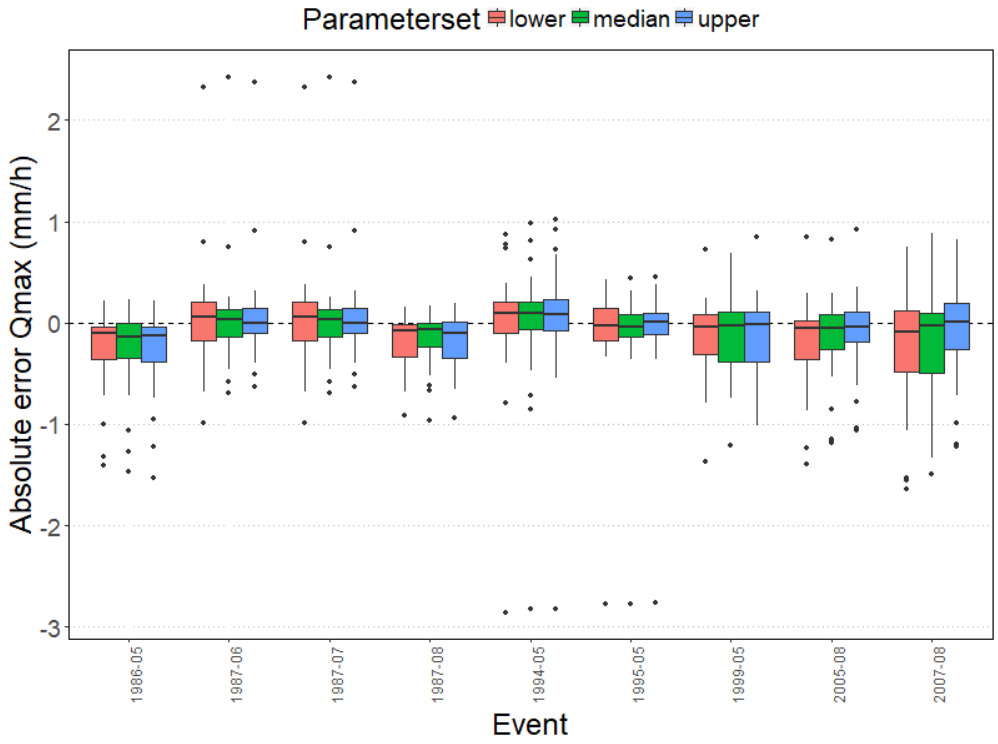


Figure 59 Comparison of the absolute volume error of the events for the different parameterizations of the HBV model.

The absolute errors were not significantly different within the events for different model parameterizations, which can be seen from boxplots and the outliers when comparing the parameterizations for the different events including all catchments (Figure 59). Note, that here some catchments are not included in all the events due to missing data.

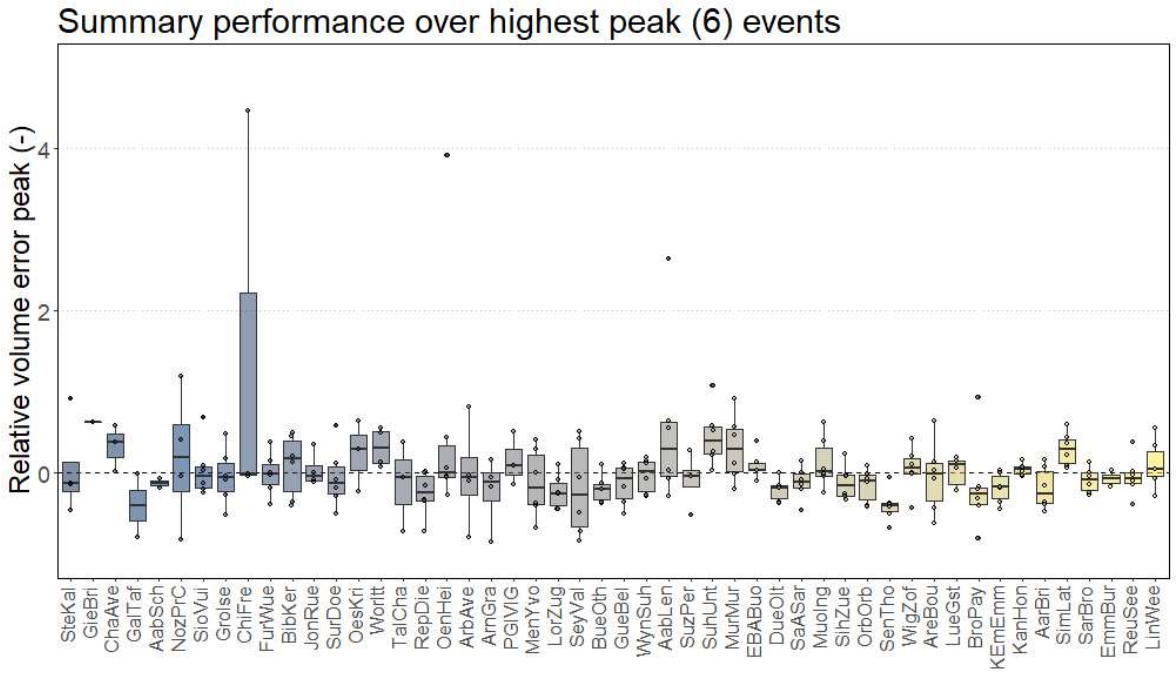


Figure 60 Relative volume error for the selected events with the largest peaks. The catchments are sorted by catchment size small (left) to large (right), the colors are according to the catchments in the map in Figure 58.

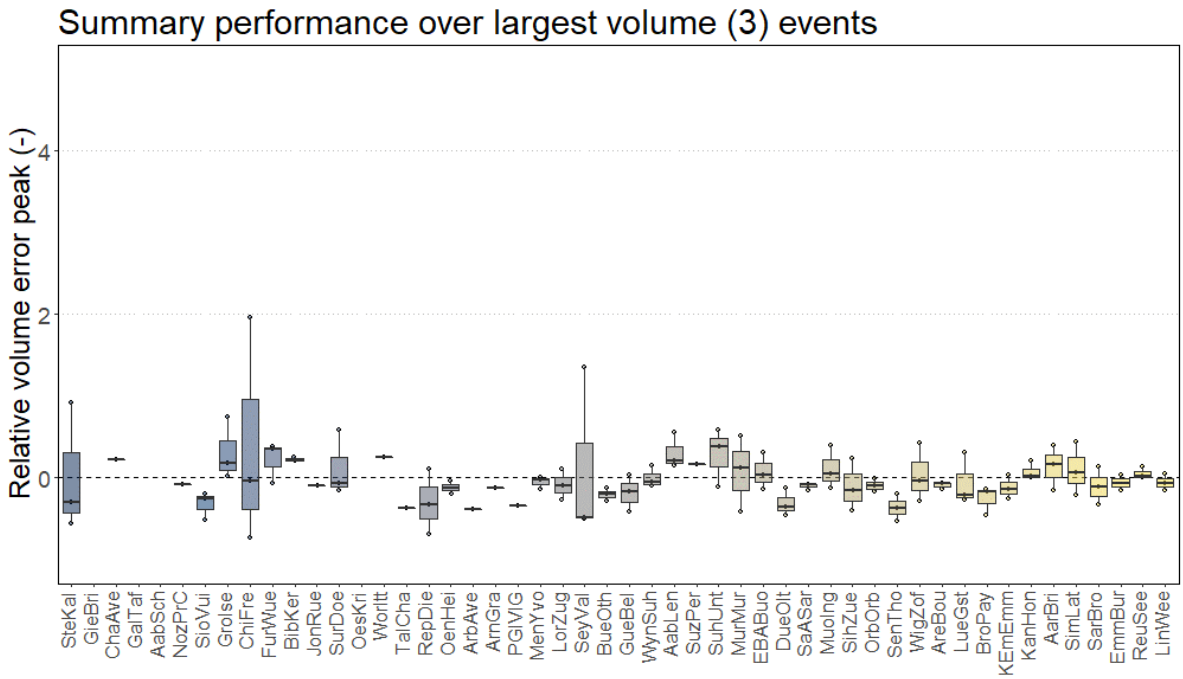


Figure 61 Relative volume error for the selected events with the largest volume. The catchments are sorted by catchment size small (left) to large (right). The colors are according to the catchments in the map in Figure 58.

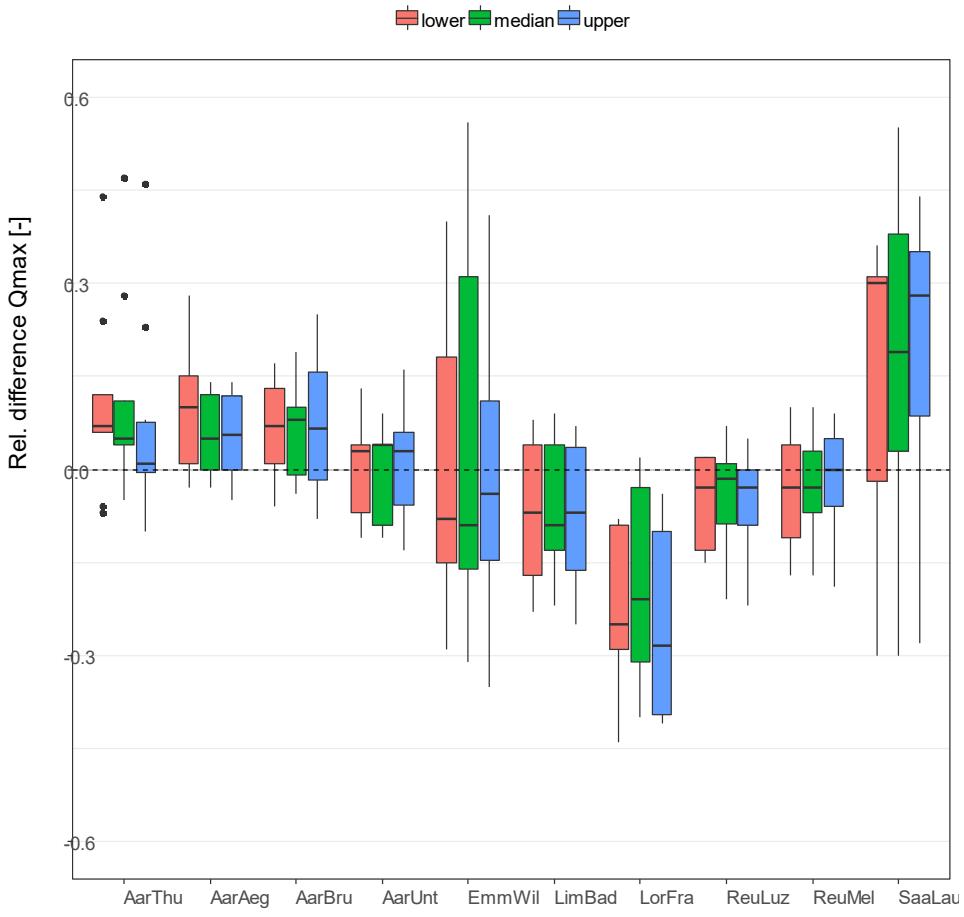


Figure 62 Difference between simulated and observed peak discharge for selected historical and large-volume events at the gauged points relative to the observed value. Different colours indicate the HBV parameter set.

For the gauge points that were simulated with RS Minerve, the observed flood peak magnitudes were matched well for the Aare, Reuss, and Limmat Rivers for most events (Appendix 10.1, Table 2).Larger deviations in peak discharge resulted for the Emme, Lorze, and Saane Rivers. The effect of the HBV parameter set aggregated over all events is low, and no parameter set seems superior.

3.7 Simulations based on GWEX generated time series

3.7.1 Spatial patterns of the top nine events GWEX

The spatial patterns of the largest events of the GWEX simulations for 24 h show different focus areas over the entire system, i.e. these largest events are driven by input from different regions. The largest event has the most precipitation falling around the Western part of Switzerland, around the Jura lakes. Most of the events have major contributions from the alpine catchments as well as the plains in the central part (Figure 63).

The spatial patterns of the largest events of the GWEX simulations for an aggregation window of 72 h (Figure 64) are less diverse than these of the aggregation window of 24 h. For many of the largest events, the spatial patterns of 72 h aggregation show strong contributions from the alpine part and in particular the eastern part of the Aare system. The runoff response to these events is more homogenous for the different precipitation inputs, however the spatial variability between the catchments in their response is very high. For the largest GWEX simulation events we found mainly the alpine catchments contributing with a large specific discharge.

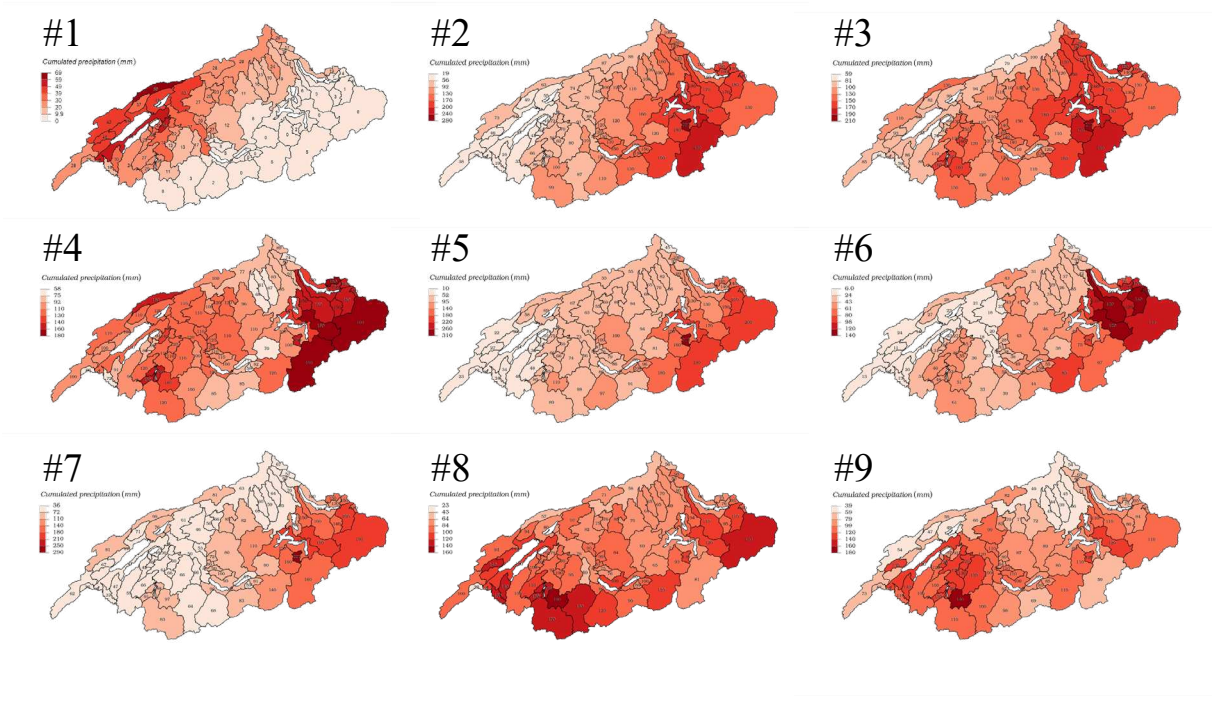


Figure 63 Precipitation patterns of the largest 9 GWEX events (24 h sums). Note that the scale at each plot is different and only the relative distribution can be compared.

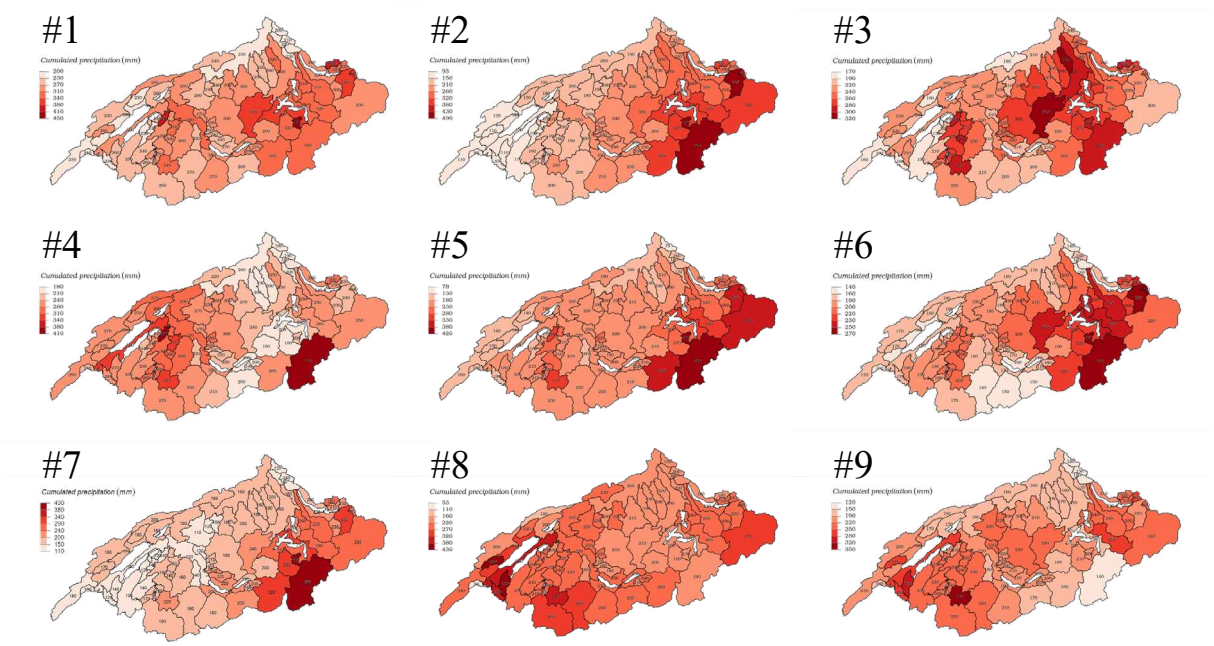


Figure 64 Precipitation patterns of the largest 9 GWEX events (72 h sums). Note that the scale at each plot is different and only the relative distribution can be compared.

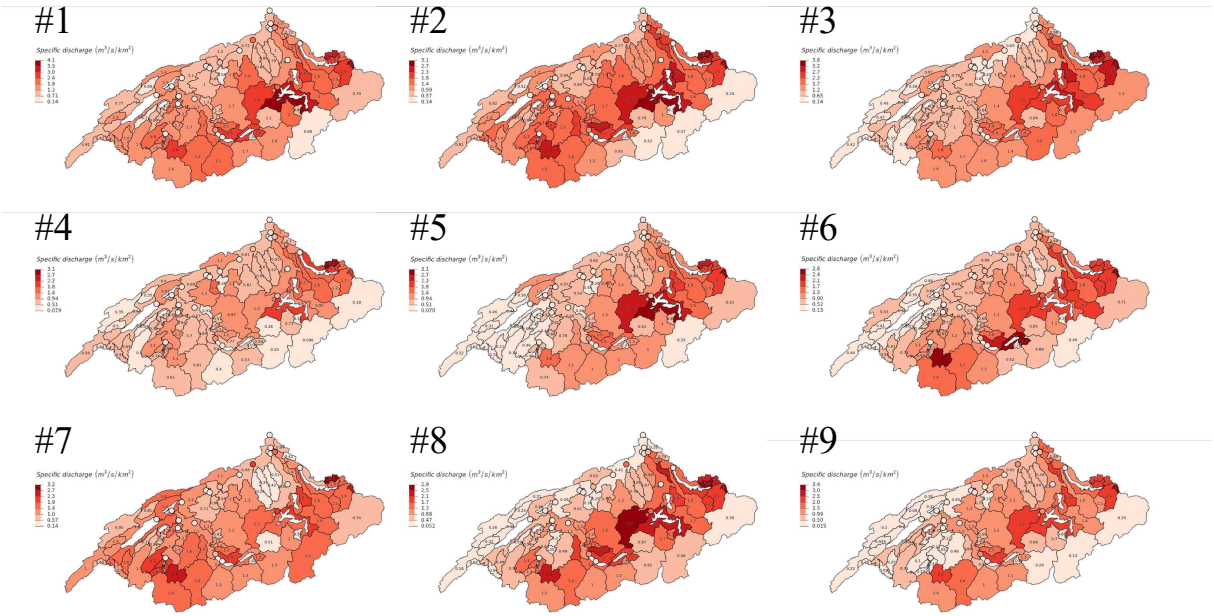


Figure 65 Specific peak discharge patterns of the largest 9 GWEX events. Note that the scale at each plot is different and only the relative distribution can be compared.

3.7.2 Plausibility of GWEX discharge distributions using FDC

The evaluation of the hydro-meteorological model chain was done at the level of each chain segment, and additionally we have evaluated the final GWEX simulations as far as feasible. For this, we tested whether the FDCs of the hydrological simulations with GWEX correspond to the FDCs of the

observations. We randomly sampled 300 periods from the GWEX simulation that have about the length of the observations that were used to calculate the observed FDC and constructed 300 GWEX FDCs. This evaluation of the plausibility of the model chain is shown for four HBV example catchments (Figure 66). The resulting GWEX FDCs are very similar to the observed ones, with the Lütschine being slightly less similar.

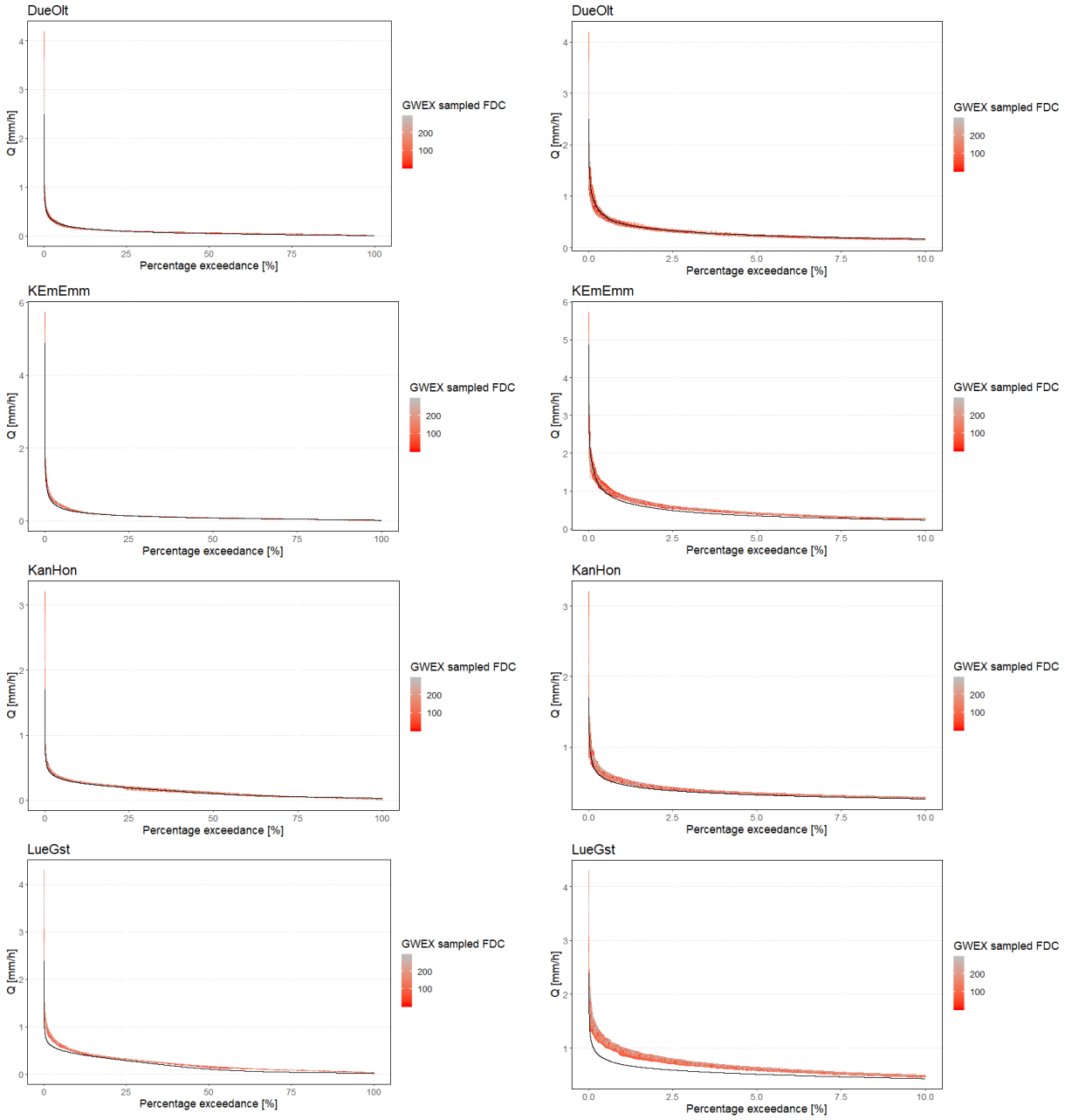


***Figure 66 Comparison of the FDCs of the a random GWEX sample* (reddish colours) *and the historical observation (black) for four* HBV example catchments (Dünnern Olten, Kander Hondrich, Kleine Emme, Lütschine). Left panel: full FDC; right panel: zoom-in to the upper 10%).**

While the plausibility of the GEWX FDCs can be evaluated for all catchments with streamflow observations, it has to be stressed that the full system and the set-up of the hydro-meteorological model chain has a focus on large catchments. In the context of this premise and the goals of EXAR, it is also there where the question whether the model chain does a good overall job is most relevant.

3.7.3 Exceedance probabilities of discharge and precipitation events

For two different aggregation windows of the discharge (72 and 144 h) the Annual Maximum Floods (AMF) as well as their preceding precipitation of 1, 3, 5 and 10 days were extracted for all years of the GWEX simulations (= 289 000 years). From the samples of AMF and preceding precipitation sums their empirical exceedance probabilities were calculated using the Weibull plotting position.

The exceedance plots have larger antecedent precipitation sums compared to the discharge for the different aggregation windows. For the lower elevation catchments when looking at discharge sums of 72 and 144 h that are smaller than 50mm the discharge is larger for some catchments compared to the antecedent precipitation. This could in most cases be explained by snow contribution to discharge which can cause the runoff ration larger than one.

Because the underlying time series is very long, the exceedance probabilities assigned to discharge and precipitation sums reached very low values (almost 1e-05), i.e. very long return periods. The exceedance probabilities of the pseudo-observations, by contrast, reached only much larger values, i.e. shorter return periods (about 1e-02) (Figure 68).

The precipitation sums were larger than the discharge sums for the large events (Figure 67). For the largest (i.e., rarest) events, increasing scatter is visible for all catchments. Comparing the exceedance probabilities from the pseudo-observations to the GWEX simulations shows that this scattering occurs at much more common events, which demonstrates the potential of the GWEX simulations to estimate rare flood events compared to the pseudo-observations (Figure 68).

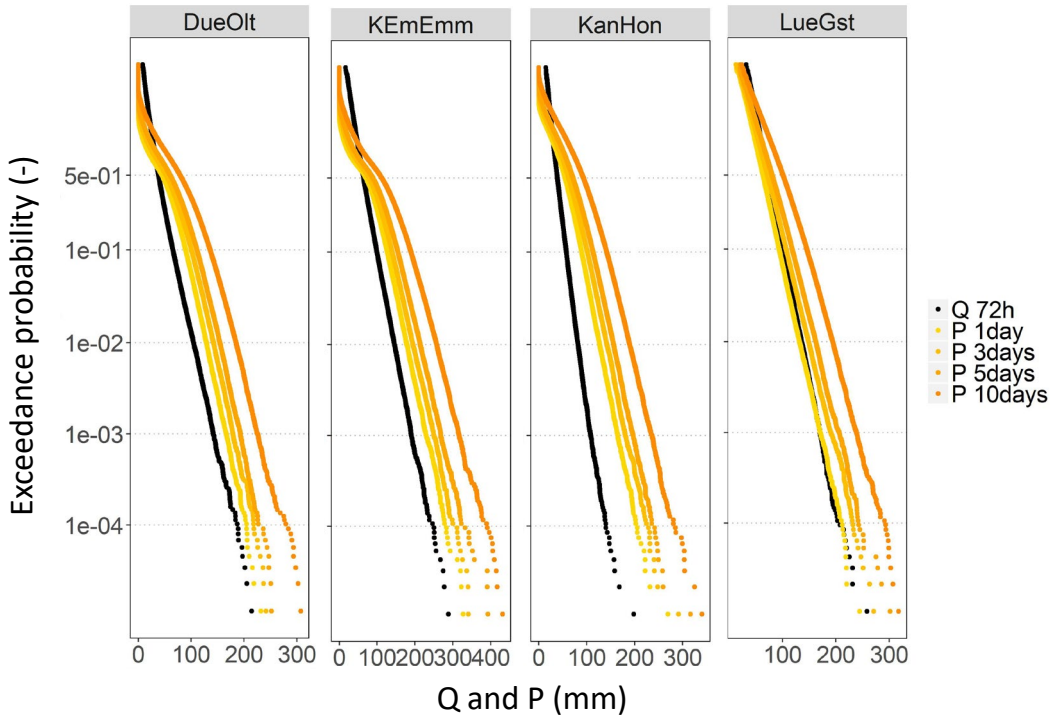


Figure 67 Examples of the exceedance plots for the catchments Dünnern (DueOlt), Kleine Emme (KEmEmm), Kander (KanHon), Lütschine (LueGst). The orange colors indicate the annual precipitation sums over different aggregation windows (1, 3, 5, 10 days), the black dots are discharge sums over 72 hours. The scale of the y-axis is the Gumbel variate (ln(ln(y)).

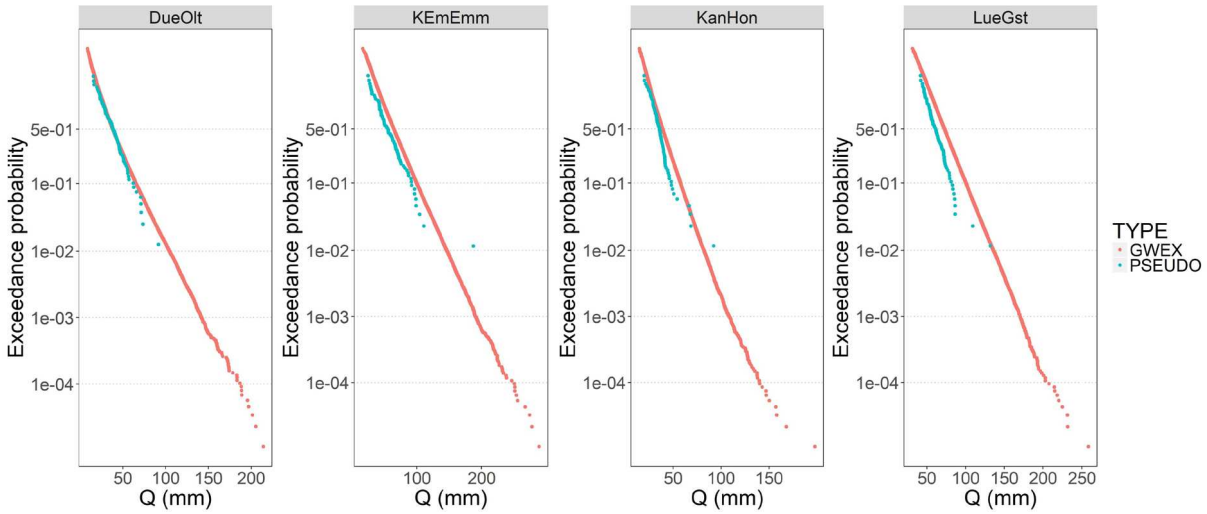


Figure 68 Comparison of the exceedances to the amounts of discharge sums for an aggregation window of 72 hours from the pseudo-observations and the GWEX simulations for the catchments Dünnern (DueOlt), Kleine Emme (KEmEmm), Kander (KanHon), and Lütschine (LueGst).

The exceedance plots from the GWEX simulations were also compared to the ones from the pseudo-observations. For some catchments the fringing of the pseudo-observations occurs at relatively large probabilities, for other catchments the shape of the exceedance curve is very similar to the ones from the GWEX simulations, though much shorter (larger probabilities) than the GWEX simulations (Figure 68).

The plausibility of the ratio between maximum event discharge sums and corresponding precipitation sums was also evaluated for all transfer points (TP). To this end, annual maximum discharge sums over 72h and 144h were identified for each TP using 289000 years of simulated discharge. Subsequently, corresponding precipitation sums were determined by aggregating areal precipitation of each TP during the event as well as 1, 3, 5, and 10 days prior to the event. Areal precipitation for each TP was aggregated from HBV catchment precipitation data.

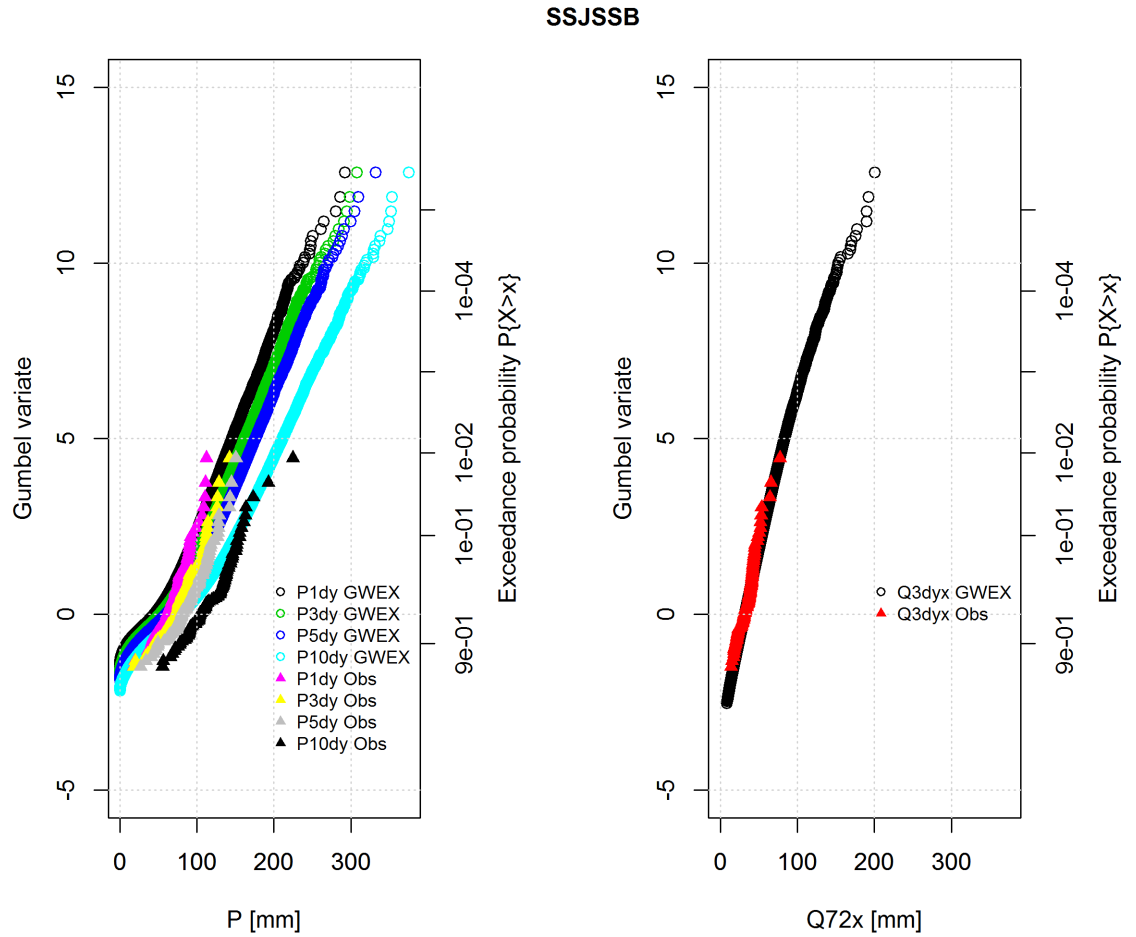


Figure 69 Example exceedance plot of annual maximum 72 hour discharge sums (right) and corresponding 1, 3, 5, and 10 day precipitation sums (left) for TP SSJSSB (Saane). Data for simulations with GWEX data (circles) as well as pseudo-observations (triangles) are shown.

Exceedance plots were made to compare the increase rate of maximum discharge sums with the increase rate of corresponding precipitation sums. Figure 69 shows an example exceedance plot for the Saane system. Data from simulations based on GWEX as well as on pseudo-observations are shown. For the GWEX data, some knees in the probability curve of aggregated discharge can be observed for low probability events (right). Similar patterns are visible in the upper tail of the accumulated precipitation curve (left). However, not all TPs show this behaviour in the probability curves of accumulated discharge (Figure 70), particularly also when they are located below large lakes that tend to level out this kind of knees with their buffering effect.

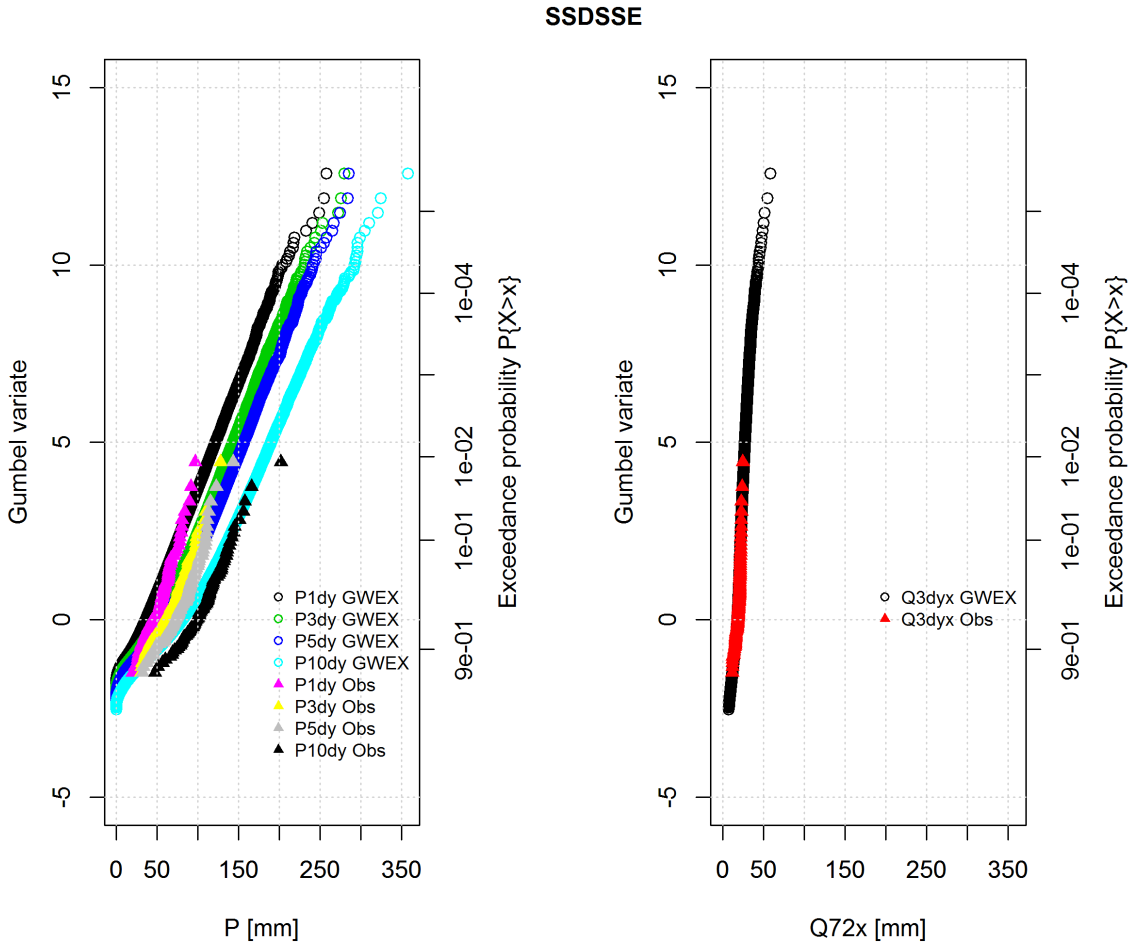


Figure 70 Same as in Figure 69 but for TP SSDSSE

Saane River and its tributaries

Comparing the annual maximum flood we found that for high return periods there is a fringing visible for some catchments. For the Saane river the tributaries and their annual maximum flood are plotted to assess where the fringing starts and from where it is derived. We found the fringing behavior already in the simulations of the HBV model for the Sarine, Sense and Jogne Rivers as well as in the Saane flood plain. However, it becomes strongest in the Sarine catchment for return periods of 50'000 years and more.

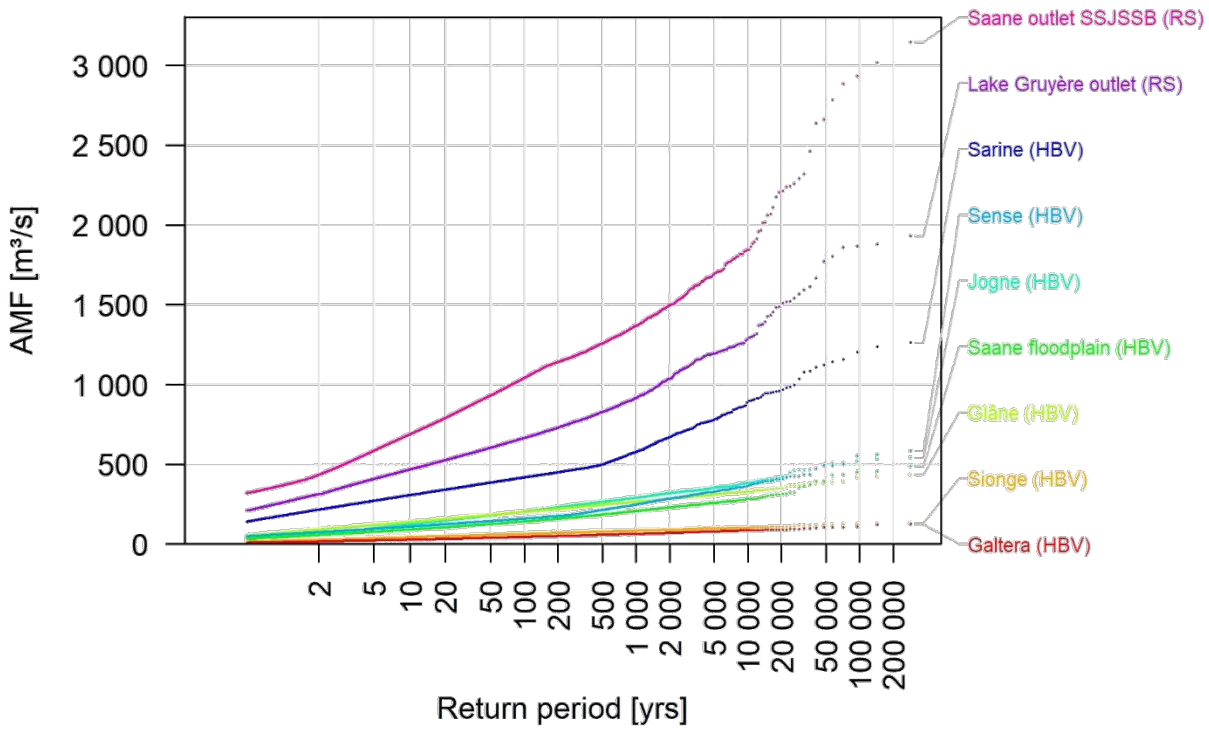


Figure 71 Tributaries to the Saane River and annual maximum floods plotted against corresponding return periods.

How catchment storage could influence this fringing behavior was assessed by looking at the HBV storage of the Sarine River (Figure 71) that has a storage that reacts very quickly to a precipitation event. These small storages can explain the fringing behavior at large return periods, because after some point the storage is filled. Any rain that is then falling on the catchment will quickly contribute to river runoff and increase peak discharge.

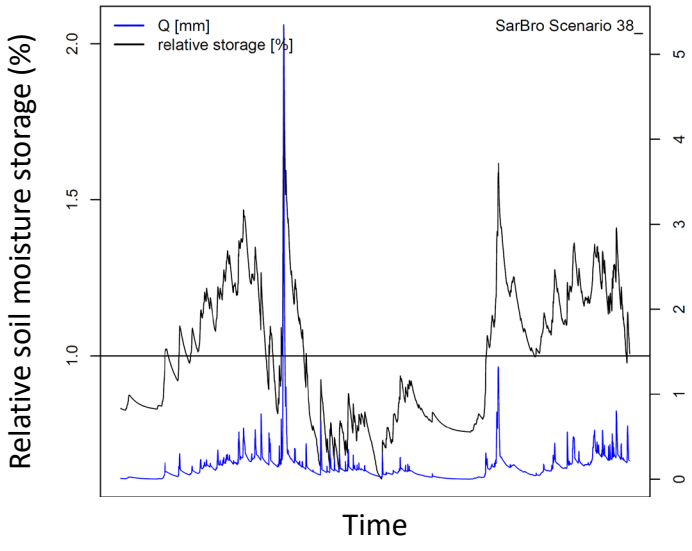


Figure 72 Comparison of relative storage state before a large discharge (Q) event. Relative storage above 1 indicates wet conditions, relative storage below one dry conditions.

3.7.4 Runoff ratios comparison GWEX and pseudo-observations

The runoff ratios were analyzed by season and for aggregation windows of 1, 3, 5 and 10 days to see if ratios higher than one can be explained by snow melt contributions (example catchments Figure 73

and Figure 74). The median of the distribution of runoff ratios is below 1 for all seasons in rainfall dominated catchments (Figure 73). However, snowmelt can be expected in March, April and May in snowmelt dominated catchments located at lower to intermediate elevations. Here, snowmelt runoff events with very little precipitation involved can lead to runoff ratios largely exceeding 1 (Figure 74). Nevertheless, runoff ratios were considerably higher than one for some events as can be seen in the outliers (black dots). Of course, larger aggregation windows show smaller runoff ratios since we compare larger sums of precipitation to the same sum of discharge.

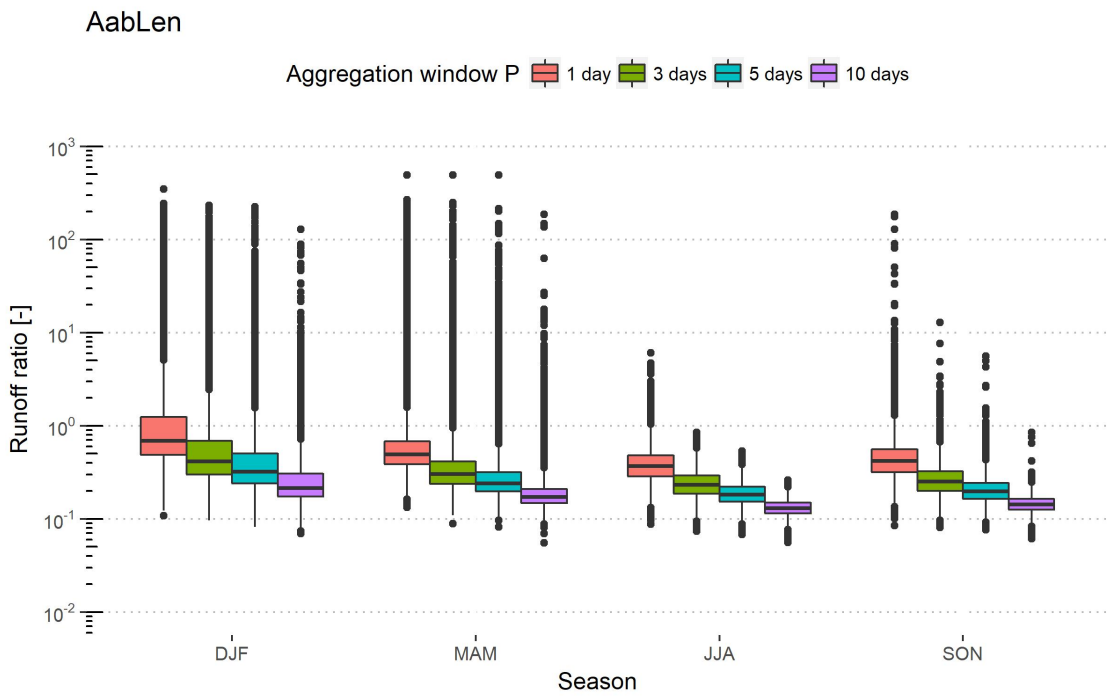


Figure 73 Runoff ratios for the Aabach catchment grouped into seasons for four different aggregation periods. Note that ratios clearly above one can occur when other sources than precipitation such as snow melt contribute to runoff.

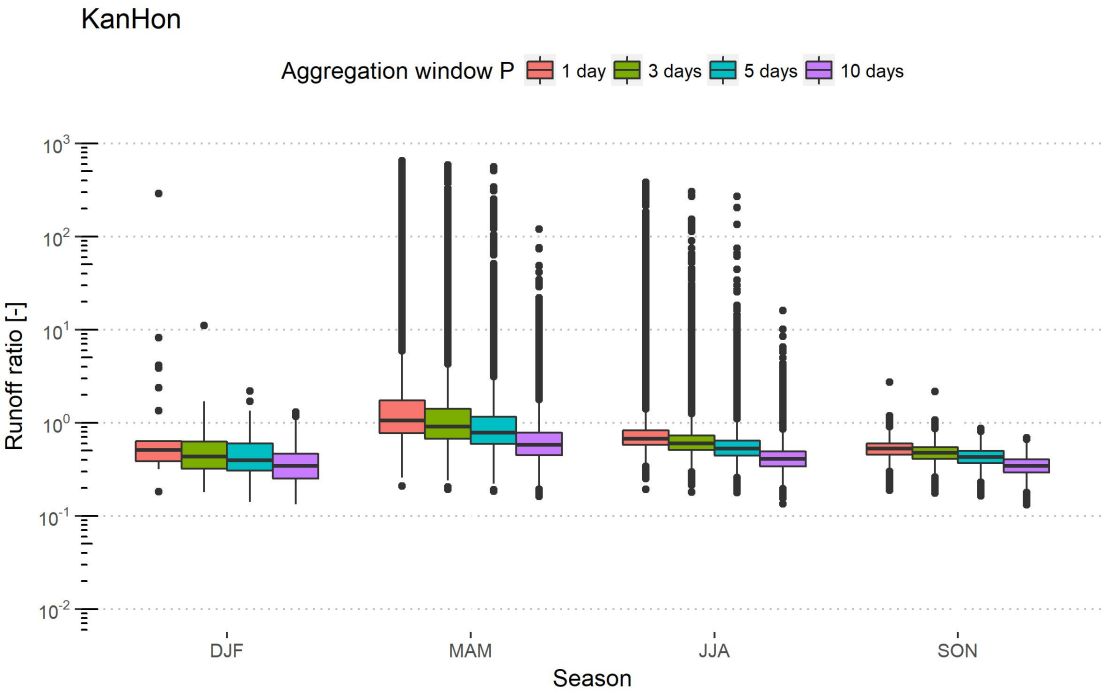


Figure 74 Runoff ratios for the Kander catchment grouped in seasons for four different aggregation periods. Note that ratios clearly above one can occur when other sources than precipitation such as snow melt contribute to runoff.

3.7.5 Initial soil moisture conditions of the largest events

Since the storage state of the catchments before the event can make a difference in the magnitude of the resulting peak discharge, we analysed the antecedent period, i.e. the initial states of the catchments before the event takes place. The initial state was derived by looking at the state of the simulated soil storage five days prior to the runoff event. This runoff was set in relation to the maximum simulated storage for this catchment to get an idea of what the relative storage filling of the catchment was before the event. The relative storage filling was then compared to the runoff ratio of the event and the 5 days prior to the event (excluding the precipitation that fell before the point where we checked the soil moisture status, i.e. prior to the five days before the event). This way it is possible to examine how much the soil moisture storage filling state was influencing the AMFs (selection: Figure 75, all catchments: Figure 76).

The relationships between relative soil moisture filling and runoff ratio are linear and positive for all catchments. This means that the more the soil storage is filled before the event, the more precipitation is turned directly into runoff. Interestingly, the soil moisture storage filling before the events varies strongly between the catchments. For instance, the large catchments of the rivers Reuss and Linth (ReuSee, LinWee) have very high storage fillings for all events. For some smaller catchments (Reppisch, RepDie; Arbogne, ArbAve), but also some larger ones (e.g., Emme, EmmBurg), the range of storage fillings before the events ranges from about 60% to more than 90%.

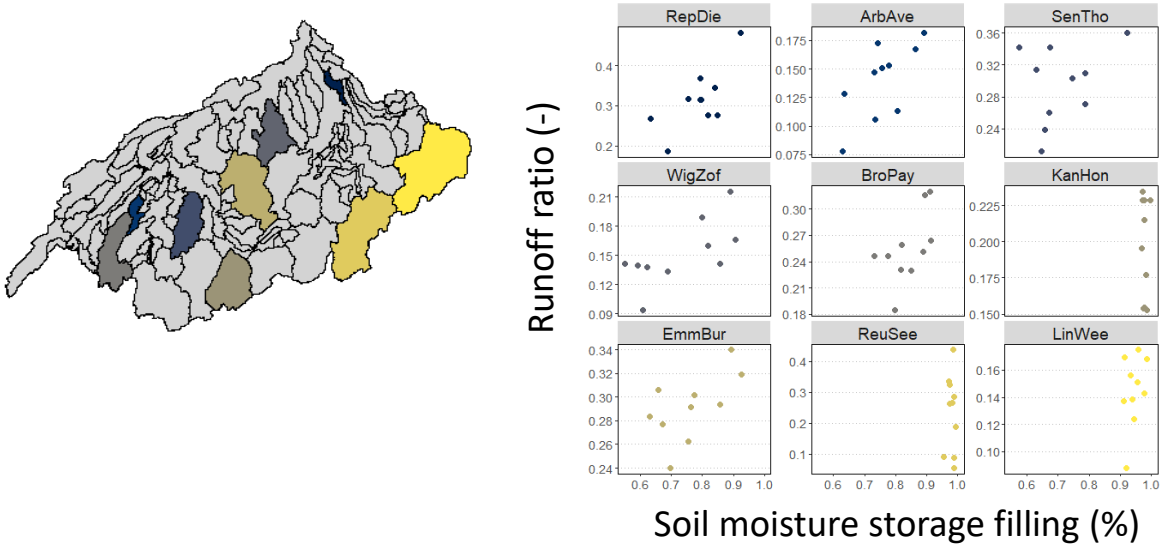


Figure 75 Relative soil moisture storage to runoff ratio for some selected catchments, sorted by size top left to bottom right. The colors of the dots correspond to the color in which the catchment is marked on the map (left panel).

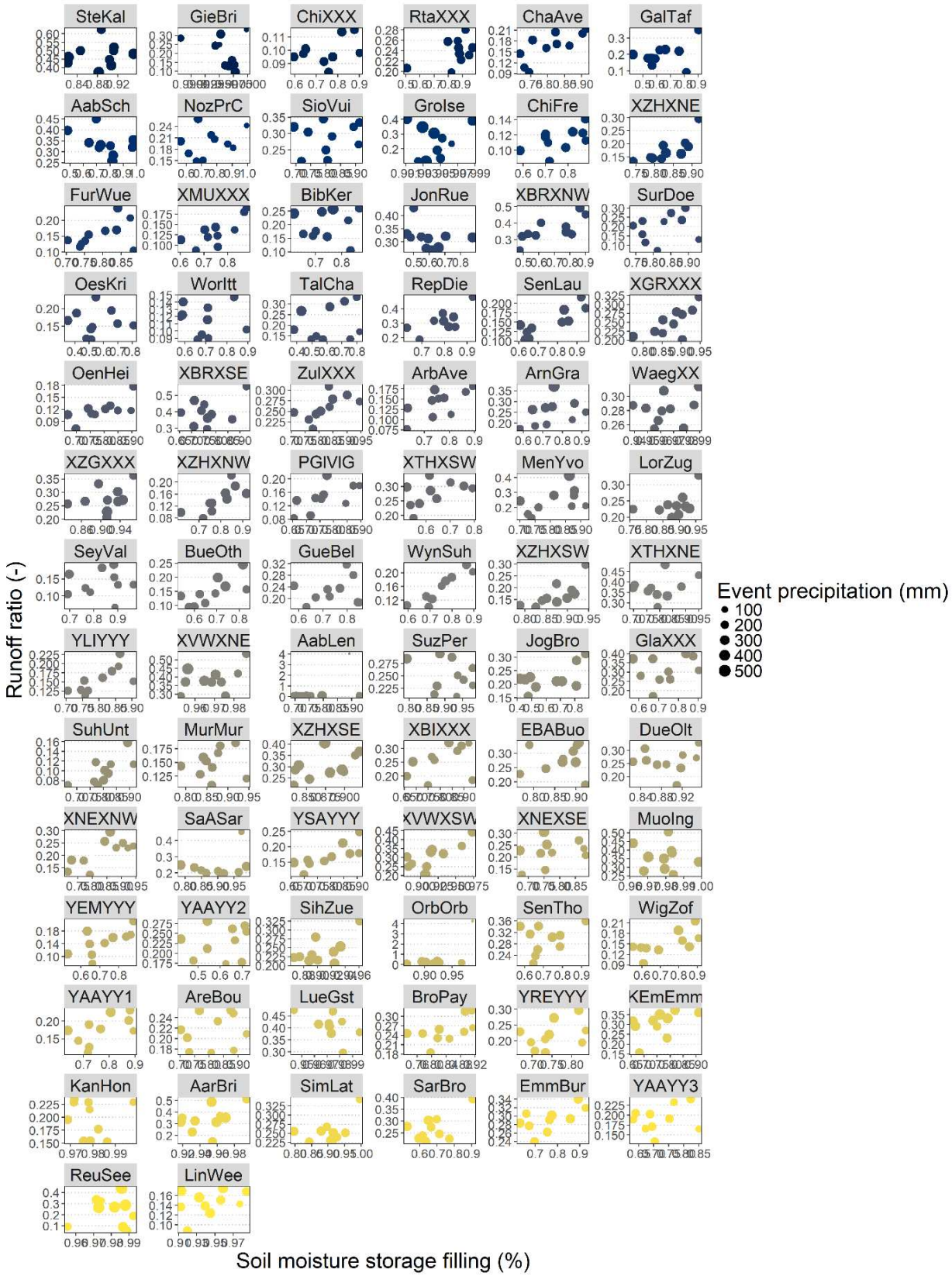


Figure 76 Relative soil moisture storage to runoff ratio for all catchments, sorted by size top left to bottom right. Color code is by catchment size as in the map in Figure 58.

3.7.6 Antecedent lake volumes

For the total system one way to assess at the storage state upstream of a certain point, as well as at the promptness to generate a flood is to look at the lake volumes before a certain flood occurs. For this purpose for the ten largest events we looked at the anomaly of lake volumes during the event year for each day of year expressed as fraction of the long-term mean value, where the long-term mean is computed averaging over the 100 years of the selected time series. Lake volumes preceding the events are generally found to be close to the average conditions, besides for data series #138 where some positive anomalies are found for the Jura Lakes and the Lake Zug (Figure 77).

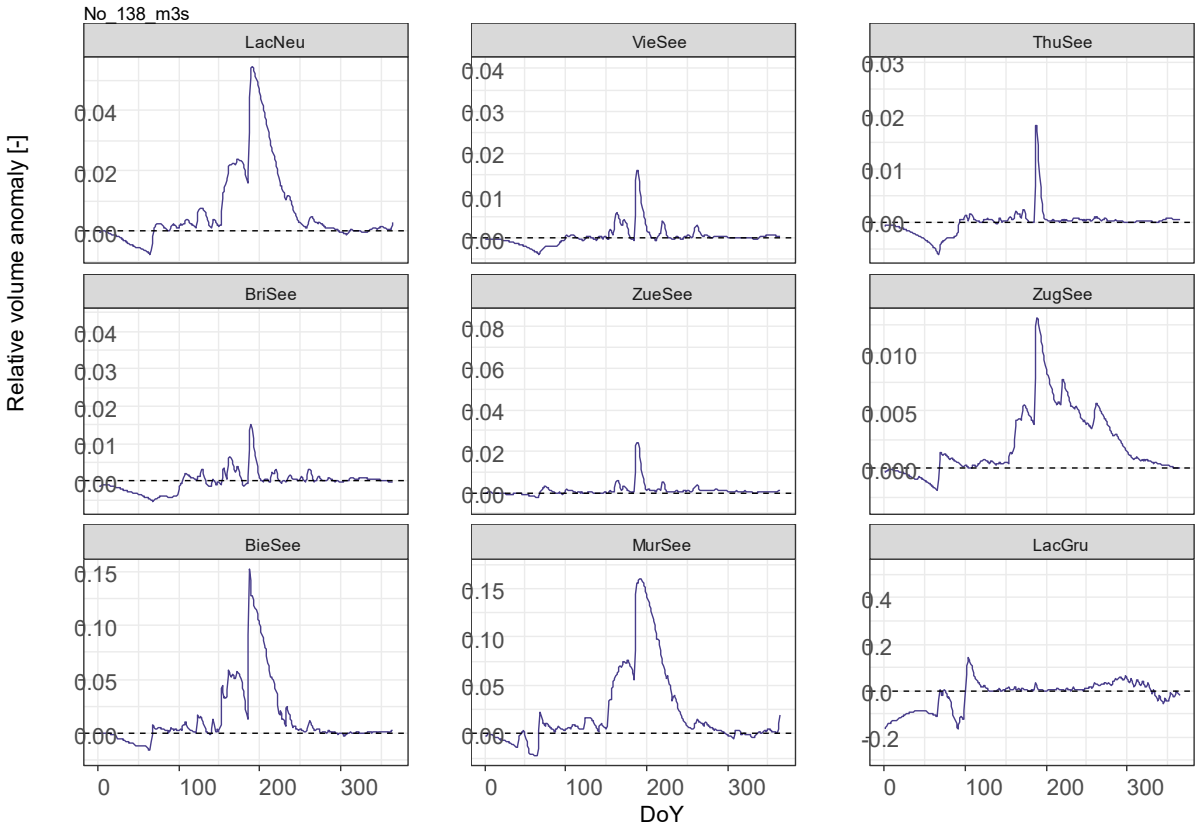


Figure 77 Antecedent lake volumes.

We further looked at the lake volume anomaly 14 days before the maximum lake volume is reached for each of the 10 events relative to the annual mean volume. For some events, pre-event lake volumes above average conditions are found, in particular for event # 138, but the derived values are rather low and plausible. The high value for Lake Murten for event # 286 is an artefact caused by the time window defined for estimating the anomaly, as for Lake Murten for this specific event a second higher peak is simulated just after the first peak.

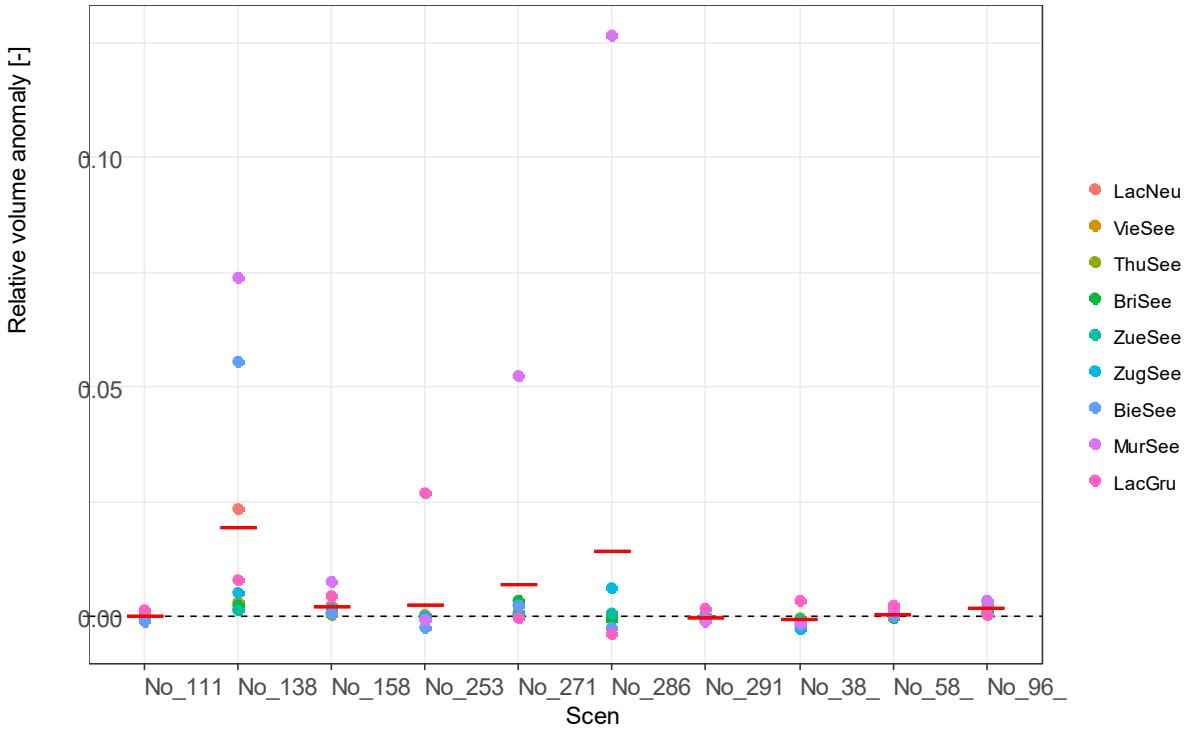


Figure 78 Lake volume anomaly 14 days before the maximum lake level for the 10 largest discharge events (x-axis) and selected lakes (colours). The red bar shows the median anomaly across lakes.

3.7.7 Snow melt contribution and antecedent precipitation during the largest events

Specific events were analyzed for different catchments in view of snow contribution and antecedent precipitation. Two different example events are discussed here. They have a different evolution in terms of antecedent precipitation and snow development (Figure 79 and Figure 80). In event #96 for the Kander (Hondrich) catchment, there is a large amount of antecedent precipitation (about 200 mm) and the wetting up of the catchment starts already about a month before the peak discharge occurs (Figure 79). Snow melt was not important as direct contribution to the event, since the zero-degree line is not raising directly before the peak discharge event. However, snow melt might have occurred at the beginning of June. In event #142 (Figure 80), accumulated precipitation is smaller (about 100 mm) than in event #96, but here snow melt can contribute directly before the event, which can be seen from the sharply rising zero-degree line before the discharge event. Event #96 has a discharge peak of 1000mm while the snowmelt event #142 has a discharge peak of only 200mm which demonstrates that large events that can occur from big precipitation events that fall on already wetted catchments.

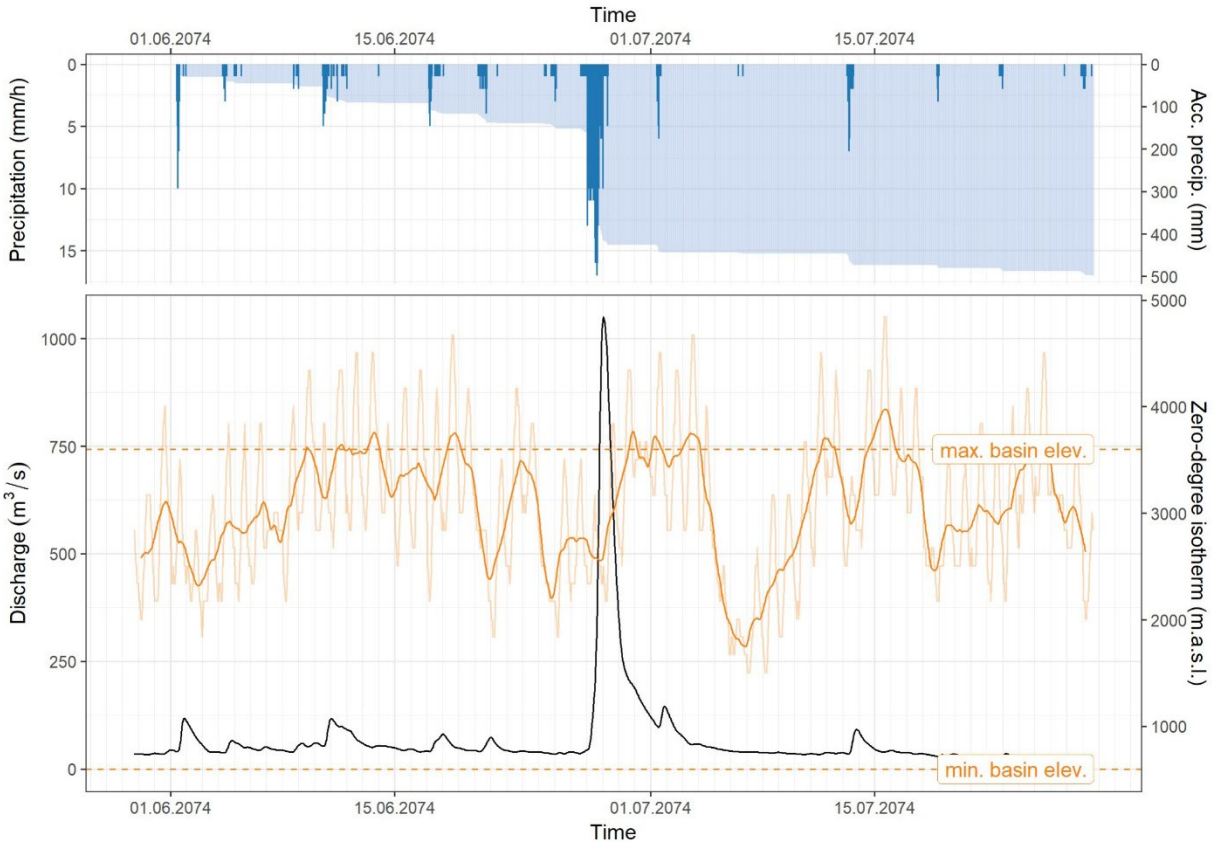


Figure 79 **Zero degree isotherm and antecedent precipitation sums for the largest event (96), in example catchment Kander (Hondrich).**

The antecedent precipitation sums vary among the largest events, meaning that in some cases the large event precipitation amounts did not fall on a thoroughly saturated area. For some catchments (e.g., Broye) we found AMFs often occurring in spring or mid-December, which points at snow melt as major driver for these floods. An example of this kind of event for the Kander catchment is shown in Figure 80. From the behaviour of the zero-degree line it is evident that snow melt occurred, while the precipitation amounts were not as large as noted for some of the major summer events.

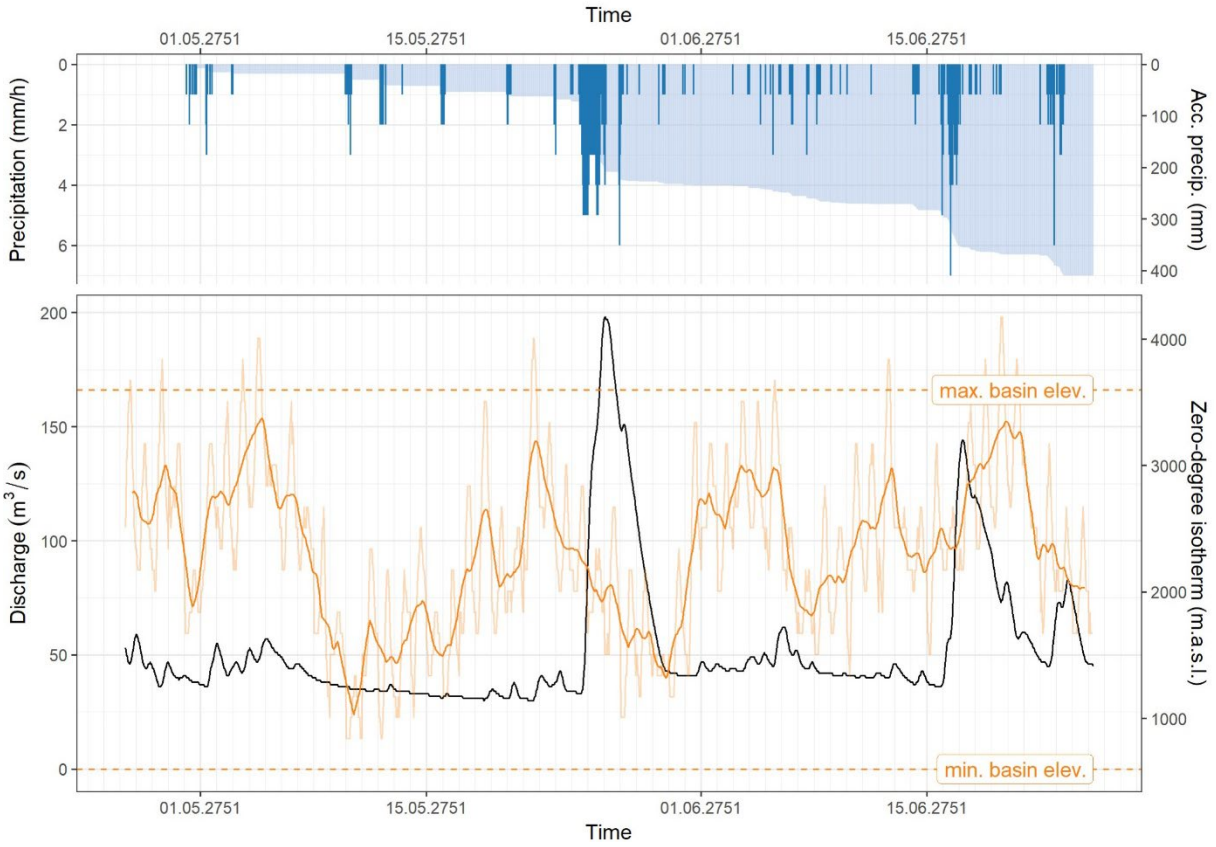


Figure 80 Zero degree isotherm and antecedent precipitation sums for a snow melt influenced runoff event (#142), in example catchment Kander (Hondrich).

To assess the realism of accumulated discharge in relation to corresponding precipitation sums on the event level, runoff ratios were calculated for each TP (Figure 81). While runoff ratios for many TPs lie in the range between 0 and 1, values above 1 occur for most TPs for selected seasons and precipitation sums. One group of such TPs are the lake outlets where runoff ratios for the 1-day precipitation sums are often larger than unity (e.g. Lake Biel, Lake Thun, Lake Lucerne, and Lake Zurich). This can be explained with the relatively longer reaction time of the lakes, i.e. longer accumulation periods are necessary to capture all precipitation contributing to the runoff event. Another reason for high runoff ratios are contributions of snowmelt to discharge in spring and possibly also in winter. Such behaviour can be observed for the Emme River (SSKSSD), and Lake Gruyère.

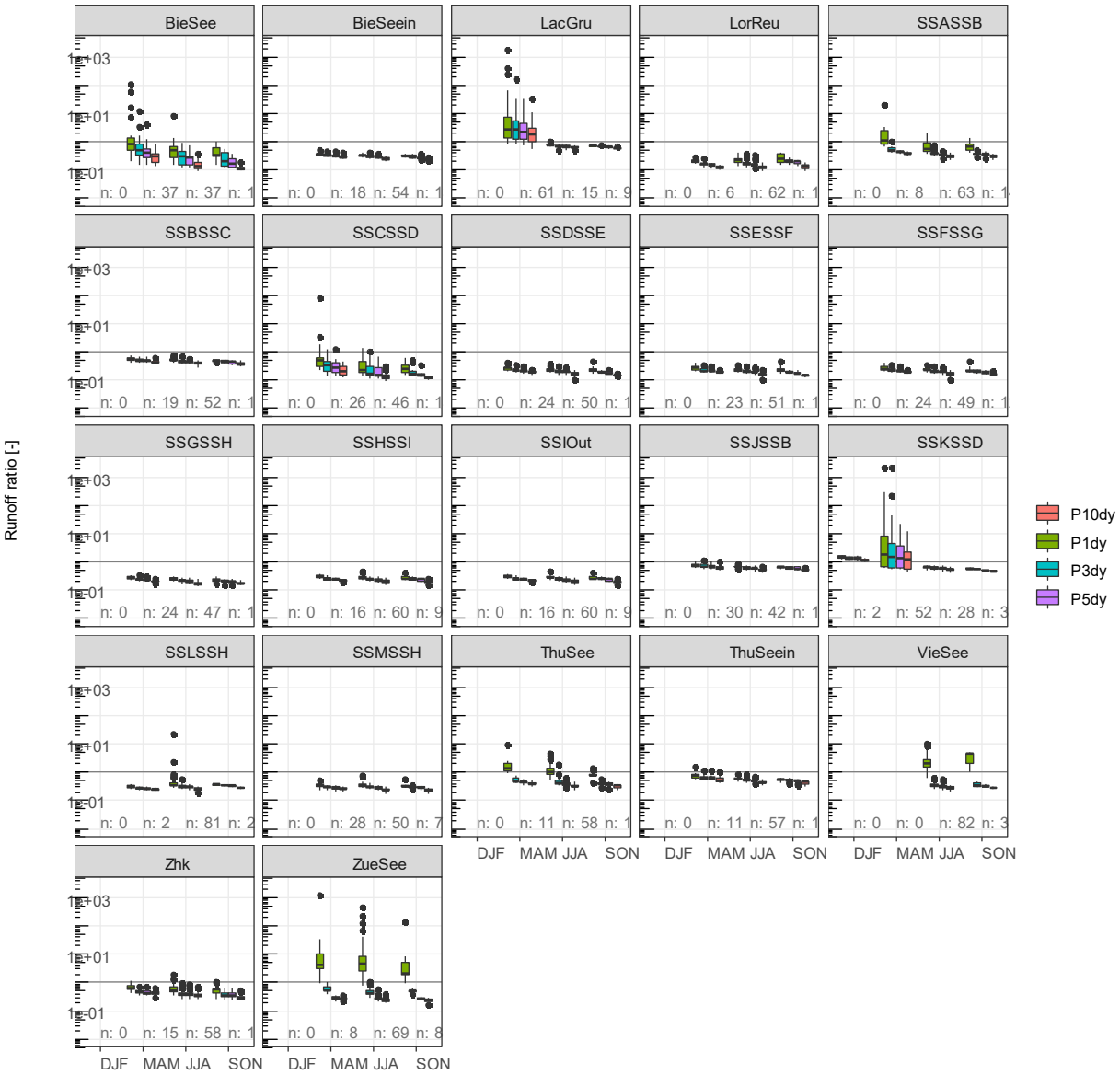


Figure 81 Runoff ratios for the 85 largest out of 289000 annual maximum discharge sums accumulated over 72 hours. Each panel shows the runoff ratios of one TP stratified by season and for the different precipitation sums (colours). Logarithmic y-axis. Horizontal line indicated a runoff ratio of 1. Note that ratios clearly above one can occur when other sources than precipitation such as snow melt contribute to runoff.

3.7.8 Plausibility of GWEX-based simulations against PMP-PMF values

In many countries, the design of dams and critical infrastructure is done with the help of the so-called PMP-PMF method. In this method, the estimated probable maximum precipitation (PMP) for a certain duration and season serves to estimate the associated probable maximum flood (PMF) (Kienzler et al. 2005, Felder & Weingartner 2017). There are several different methods that can be used to do this but there is no scientific consensus about a method to be preferred. Because of the large uncertainties that are inherent in PMP and PMF estimations, it is generally recommended to compare the results with estimations from other methods and thus to evaluate their plausibility. The WMO (2009) states that "it should not be a requirement that PMP/PMF should be larger than or smaller than a storm/flood with a defined frequency, as long as the estimations are undertaken in a reasonable manner ". In recent publications the PMP/PMF method was applied with varying degrees of complexity to Swiss catchments of different size and characteristics, e.g. to the Aare River catchment down to Bern (Felder

& Weingartner, 2016; Felder & Weingartner,2017; Zischg et al. ,2018), to the Kander River at Hondrich and to the Emme River at Wiler (Felder et al. ,2019) as well as to the Sihl River down to Zürich (Kienzler et al., 2015). As a first step, the PMPs were estimated considering physical plausibility criteria as well as relationships between depth, area, and duration of precipitation. In a second step, these were re-distributed in time and space, either deterministically with the help of prescribed simplified distribution patterns (Kienzler et al.2015), or probabilistically using a Monte-Carlo approach (Felder & Weingartner, 2016; Felder & Weingartner, 2017; Zischg et al., 2018; Felder et al., 2019). In all mentioned studies PREVAH was used (Viviroli et al., 2009) as hydrological model. In addition, a 1D hydrodynamical model was coupled with PREVAH in Felder & Weingartner (2017), Zischg et al. (2018) and Felder et al. (2019) to account for effects of overbank flow and to assure a better routing. It is possible to put the simulations done in EXAR into context by comparing them to the studies mentioned (Table 4).

Table 4 Overview of PMF-values from literature and the simulations based on the outputs of the weather generator GWEX.

River, location	Study	TP or catchment	PMF [m^3/s]	Q_{max} GWEX [m^3/s]
Aare, Bern	Felder&Weingartner2016,2017, Zischg et al.2018	SSASSB	1296	1250
Emme, Wiler	Felder et al.2019	SSKSSD	1388	1356
Kander, Hondrich	Felder et al. 2019	KanHon	830	1050
Sihl, Zürich	Kienzler et al.2015	SihZue	975	772

The highest peak values resulting from simulations with GWEX input are either very similar or in the same order of magnitude as the ones reported in the literature. The largest deviations are found in smaller catchments such as the Kander or Sihl Rivers, in which the variability of the distribution of precipitation is more relevant for runoff generation. Overall, it can thus be concluded that the simulated maximum peak values in EXAR seem plausible. Given the different methods applied and the uncertainties inherent in each of these methods, this strengthens the confidence in the EXAR results.

3.8 Comparison of simulation from GWEX and SCAMP for the 10 largest events

To examine whether using a different weather generator would lead to different runoff peaks, the analog-type weather generator SCAMP was used as an input to HBV. Simulations were done with over 3000 of SCAMP's largest precipitation events, and the largest 10 resulting runoff events were compared with the largest 10 runoff events produced with the standard weather generator, GWEX.

3.8.1 Annual maximum floods

Comparing the discharge sums over 72 hours derived from the two different weather generators (GWEX and SCAMP), we found that the SCAMP weather generator results in higher amounts, particularly for smaller catchments (Figure 82). For the larger catchments this difference became smaller, however also here it was apparent. For some catchments there is a greater difference between the two weather generator simulations. Only for very few catchments the GWEX simulations resulted in higher 3-day discharge sums than the SCAMP simulations, namely Aabach (AabLen), Sionge (SioVui), Worble (WorItt), Lake Gruyère tributaries (XGRXXX), Jogne (JogBro), Sarner Aa (SaASar), Sarine (SarBro) and Emme-Burgdorf (EmmBur), of which most are among the smaller catchments.

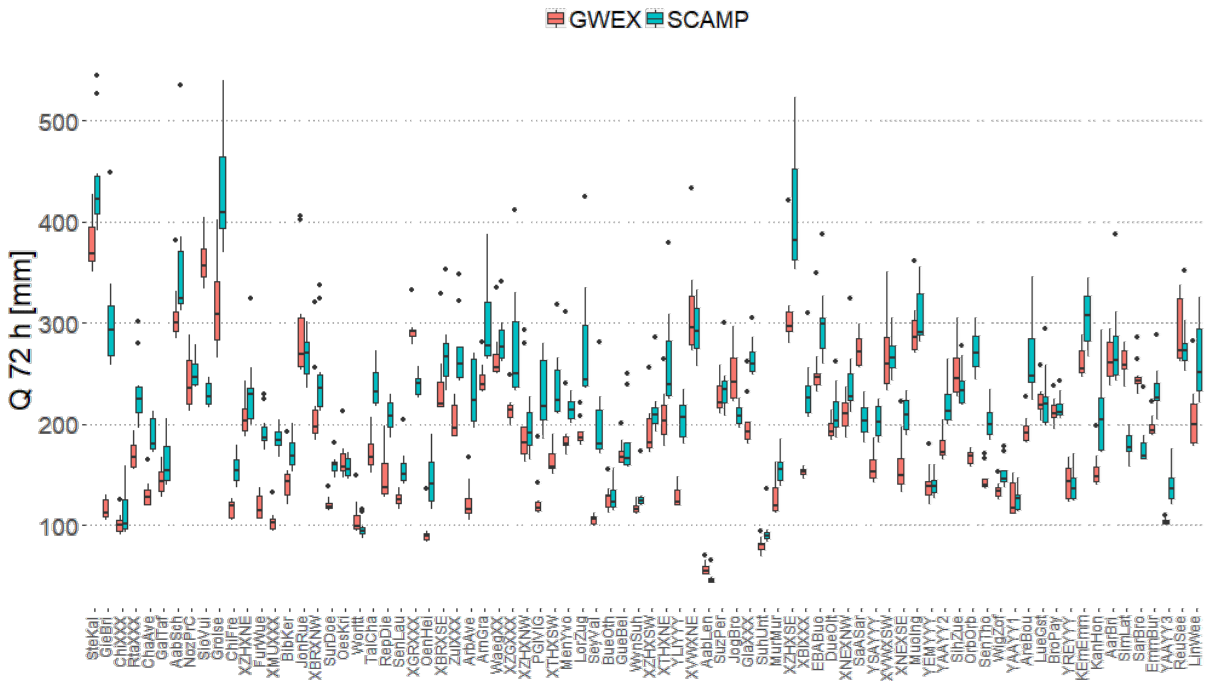


Figure 82 Comparison of the 3-days-sums of the ten largest events for the GWEX and the alternative SCAMP weather generator based simulations.

In order to compare GWEX and SCAMP simulations for the total system we extracted the ten events that lead to largest AMFs at the outlet. The two weather generators lead to comparable results: for many TPs the largest events are higher for SCAMP, whereas the median is similar or even higher for GWEX. In a few cases GWEX leads to the highest values, namely for the outlet of three lakes at ThuSee, VieSee, ZueSee. This is partly due to the interplay of differences in the seasonality (see next chapter) and the monthly lake regulations, and is overturned at the following transfer points depending on the influence of the tributaries downstream.

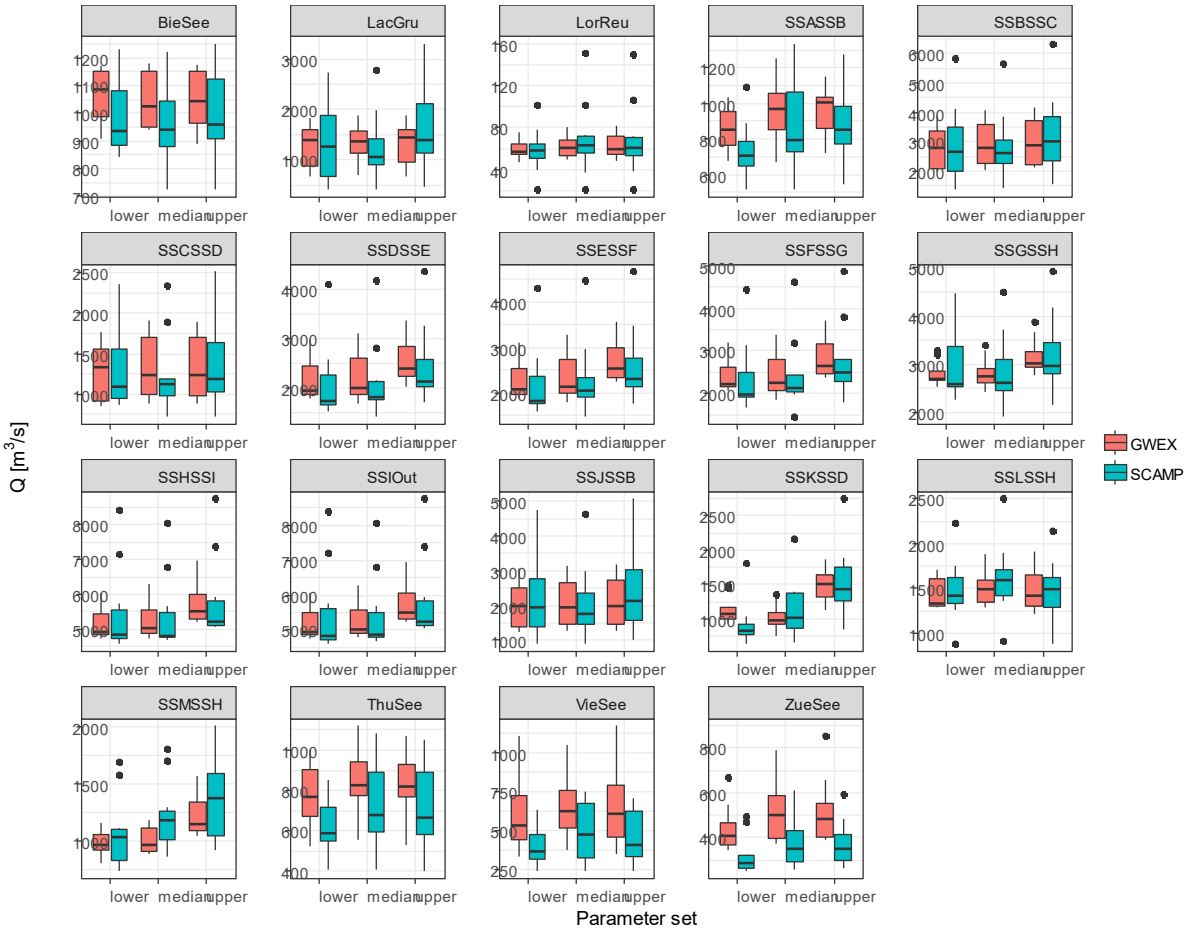


Figure 83 Flood discharge simulated with GWEX at all TPs for the ten largest events at the outlet, for each of the three parameter sets (red), compared to simulations performed with the alternative weather generator SCAMP (blue).

3.8.2 Seasonality

The ten largest discharge events produced with GWEX input occur in the late spring or summer seasons, i.e. between May and September (Figure 85). Thus, snow-melt processes do not play a major role in the flood generation in most of the sub-catchments for these largest events, even though in May snowmelt might be still possible. Snow processes still could be of local importance in headwater sub-catchments located at high elevations with snow or glacier covers, but these processes become less relevant at the larger scale of the entire Aare River.

For the SCAMP simulations we found that the 10 largest discharge events occurred also in the summer months. However, interpretation has to be done carefully here since the first selection of GWEX events to run through HBV already favoured the selection of summer events. This selection bias might be responsible for the occurrence of summer AMFs.

The seasonality of the then largest events defined by the 72 hour sums of discharge are the same for many catchments comparing GWEX and SCAMP simulations; for others there is a distinct difference between the two (Figure 84). We have to keep in mind again that the events for the alternative weather generator (SCAMP) were selected based on precipitation. Therefore, the sample of events is already biased to heavy rainfall events, and floods with significant contribution of snow melt are largely removed. In the GWEX simulations, however, the 10 largest floods were selected from the entire simulation period of 289 000 years, meaning that snowmelt floods are included as well.

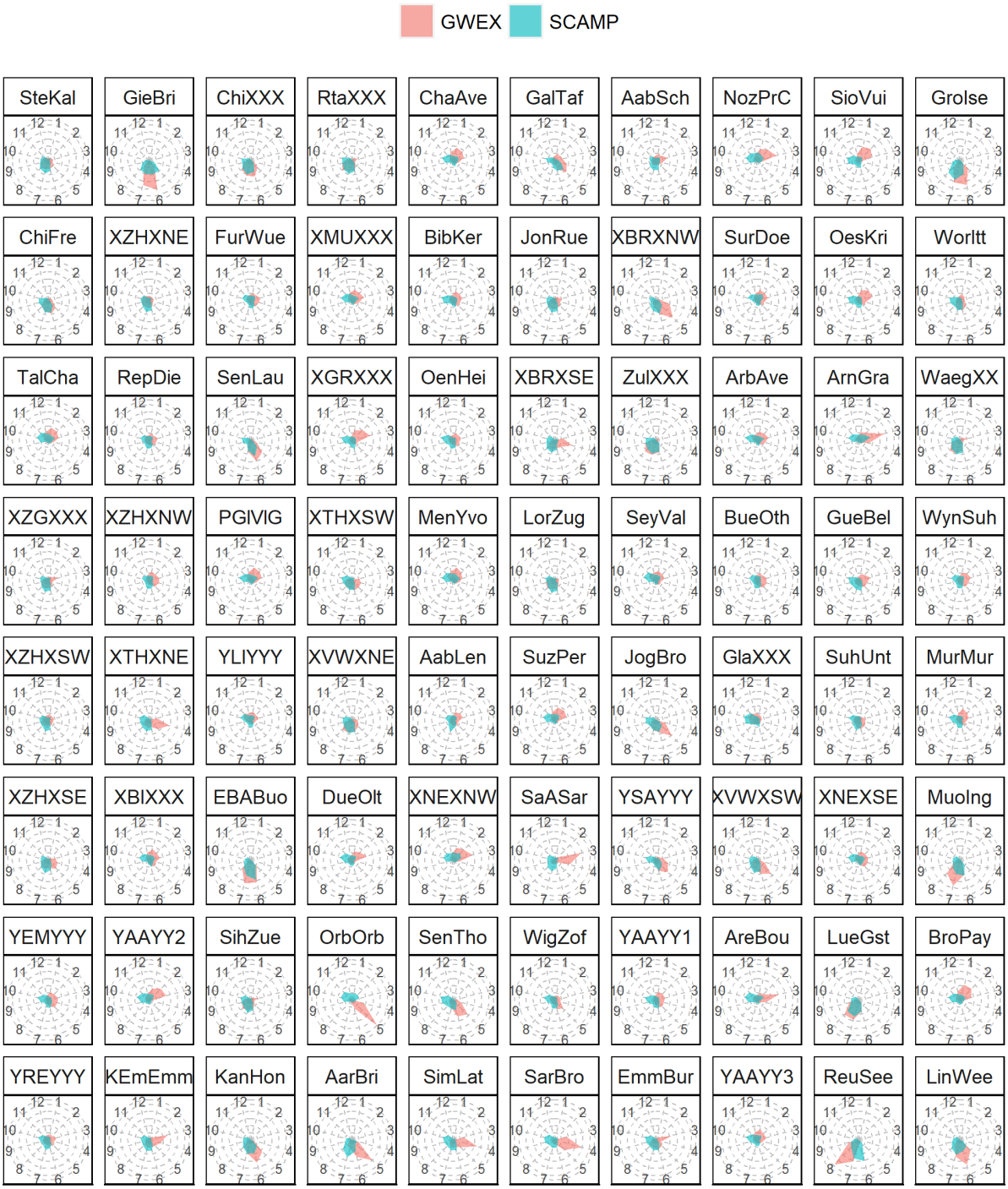


Figure 84 Seasonality spider plots comparing the AMF discharge sums over 72 hours. The circles count the events that occur in a certain month. The months are indicated with the numbers 1 to 12.

Catchments with a nival flow regime (Figure 85) appear to have a strong sampling bias towards the summer months for the SCAMP simulations, while we found the expected snow melt floods for the GWEX simulations.

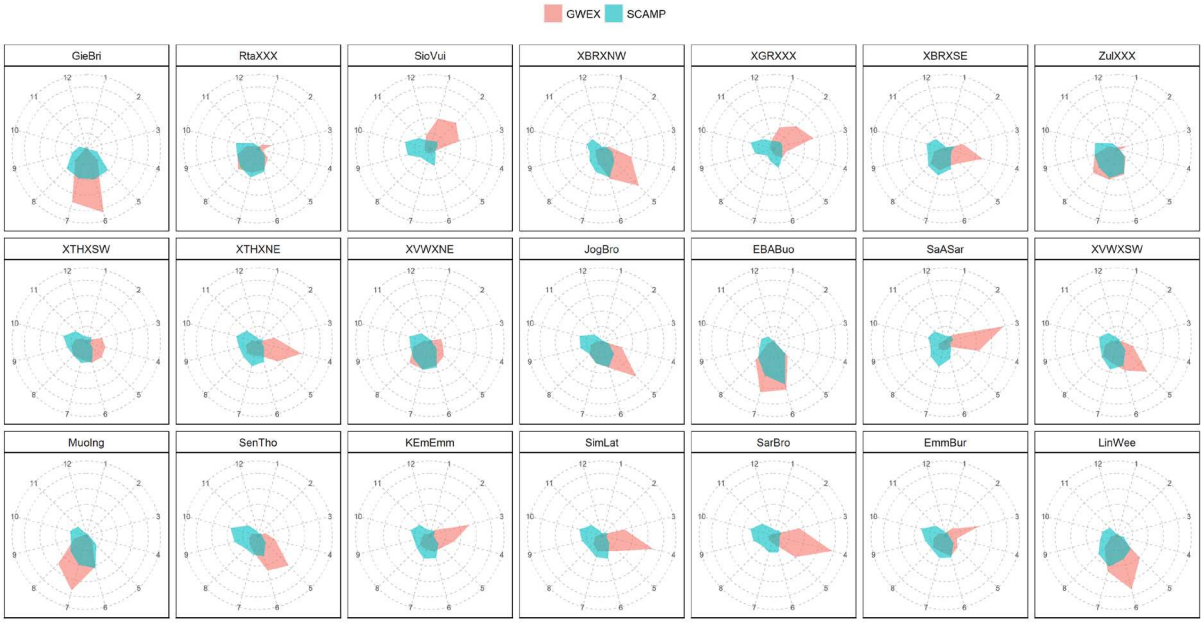


Figure 85 Comparison of the seasonality between GWEX and SCAMP simulations for catchments with a nival flow regime.

As we already did before for the pseudo-observations and the GWEX simulations, we looked at the seasonal distribution of all available AMF for all TPs of the total system, which for SCAMP amounts to 3425 events in total (Figure 86). Again here applies the same as for the hydrological simulations: we must pay attention at the sample size, which for SCAMP is significantly lower than for GWEX, as well as at the event extraction method. Generally we see a more uniform and spread distribution with the alternative weather generator SCAMP, also reflected in a medium seasonality (between 0.42 and 0.64) and a rather similar mean date of AMF occurred across the TPs, occurring mostly in August or early September, i.e. later compared to GWEX.

The same characteristics become even more evident extracting the month of occurrence of the ten largest events at the outlet of the catchment (Figure 87).

3.8.3 Conclusions

Overall, the GWEX and the SCAMP weather generators lead to simulations that are comparable in seasonality, keeping the sample bias in mind. The SCAMP simulations resulted in AMFs in a similar range as the GWEX for most catchments, with a tendency to higher AMFs for most catchments.

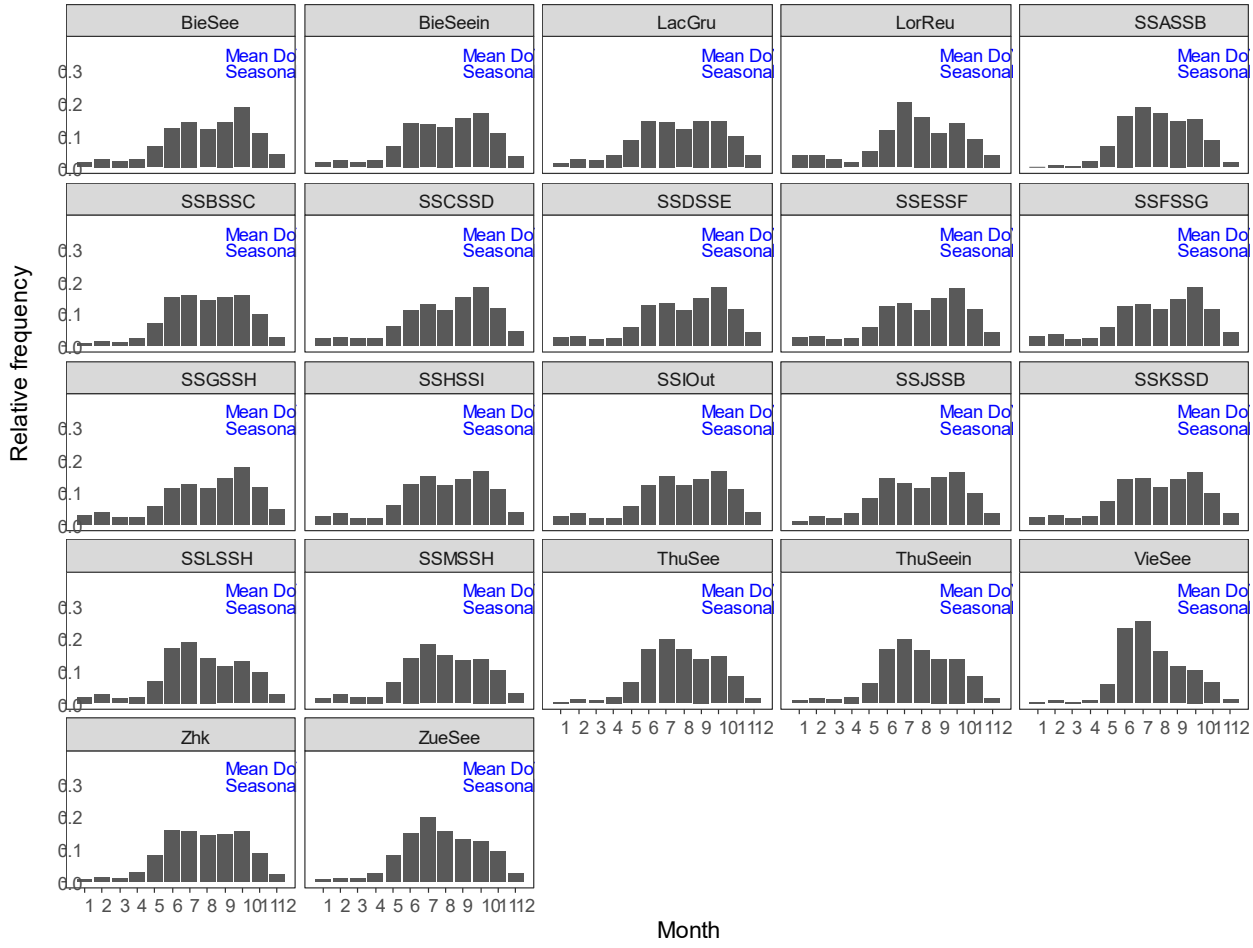


Figure 86 Histograms of the month of AMF occurrence of the SCAMP 3425 simulated events. Numbers in blue are circular summary statistics indicating the mean date of occurrence across all years and the strength of seasonality.

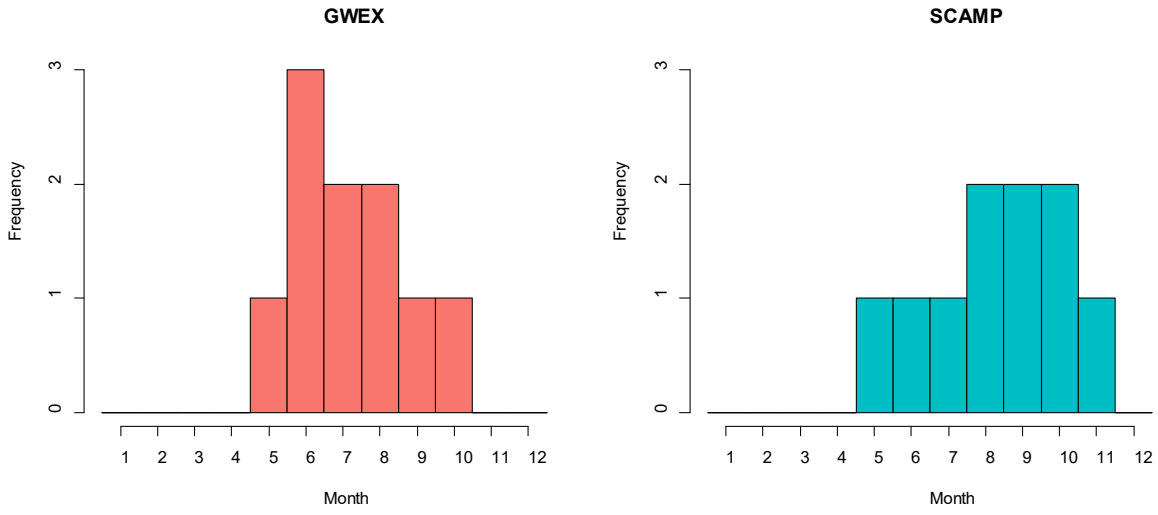


Figure 87 Frequency of the month of occurrence of the ten largest events at the outlet of the catchment for GWEX (left, red) and SCAMP (blue, right).

4 Correction of inconsistency in AP2 hydrological model runs

Daniel Viviroli, Calvin Whealton

In fall 2018, an anomaly was discovered in a small fraction (3.5%) of the hydrological simulations of AP2. In detail, 11 out of the 300 scenario blocks with a length of 1 000 years each showed lower values for annual maximum floods (AMFs) than the other 289 scenario blocks (Figure 88). This anomaly was not visible in the results of the individual catchments as simulated by HBV, but only for sites at the Aare River as simulated by the total system runs in RS Minerve. An analysis with data resampled from the 289 scenario blocks confirmed that it is highly unlikely that the deviation observed in these 11 blocks is due to stochastic variability. After intensive search, the source of the anomaly could be narrowed down to the HBV runs of scenario blocks 2–12. Most likely, the problem was caused by a file transfer problem, i.e., a silent crash of a copy process. In consequence, the simulations for the affected scenario blocks 2–12 used as input the results from an outdated version of the weather generator (G-WEX), while the G-WEX results on AP2's local storage were indeed the correct, most recent version.

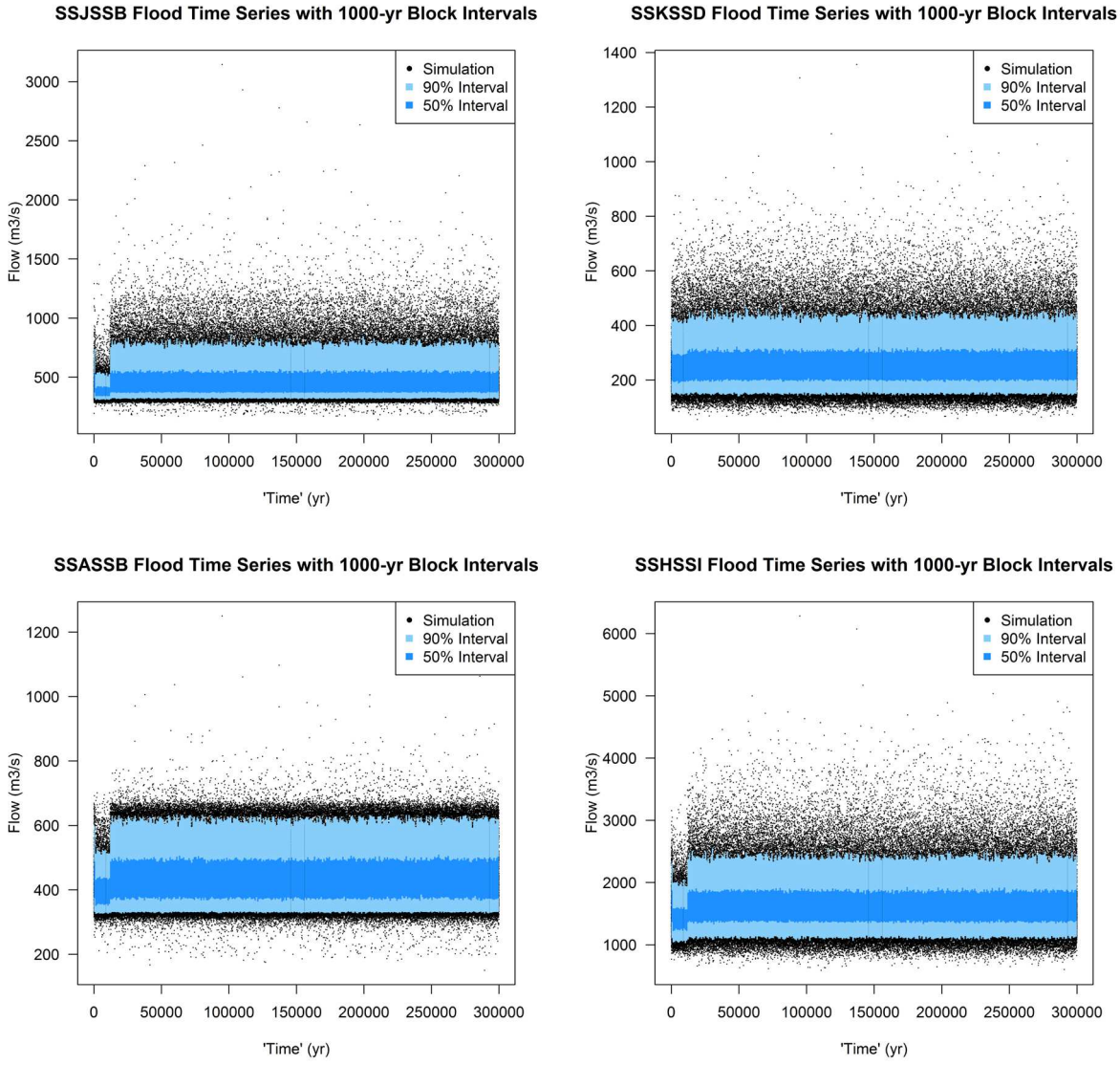


Figure 88 Anomaly found in simulations, shown at four selected sites (SSJSSB = Saane River, SSKSSD = Emme River, SSASSB = Aare River at Halen, SSHSSI = Aare River at Untersiggenthal, Stilli). The plots show all simulated annual maximum floods (AMFs), and the 50% and 90% intervals of the AMFs for blocks of 1000 years.

AP2 then ran a re-simulation of blocks 2–12 with the most recent G-WEX results in HBV and subsequently also in RS Minerve. Results were available in January 2019, and it became evident that the AMFs of blocks 2–12 are still not statistically consistent to that of the other 289 blocks, although differences were much smaller. The most marked deviating behaviour was found at SSJSSB (Saane River, Figure 89). Further minor differences were identified visually in the Reuss and Limmat Rivers, and statistically at other transfer points, many of which are likely influenced by the upstream differences. The cause for this inconsistency was identified in details of the simulation set-up (model parameters) that was not fully identical in the original simulations made 2018 and the re-simulations of 2019. A large part of the inconsistency can be explained by an update made for the parameter set in HBV's representation of the Sense River catchment (share in total catchment area of Aare River ~2.5%). For this catchment, the BETA model parameter was changed after the 2018 simulations because it was set at its lower physically plausible limit. BETA defines the relative contribution of rain and snow melt to runoff. The simulations for the Sense River were not updated after the 2018 model runs, however, because these runs were made ahead of schedule to furnish APs 3 and 4 with preliminary results. However, because there appeared to be no major issues in these preliminary runs, these served as final results, saving 2–3 months of computing time, and leaving more time for scrutinising the extreme values found. The updated parameters for the Sense River thus did not come into effect until AP2's re-simulation of 11 blocks in 2019, where they became main responsible for the emerging inconsistency. Another difference emerged in RS Minerve via a similar course of events: There, minor adjustments were made after the 2018 simulations with the goal of improving simulation stability of RS in the Reuss River subsystem. These adjustments have impact on SSLSSH (Reuss River), SSHSSI (Aare River at Untersiggenthal, Stilli) and SSIOut (confluence of Aare River with Rhine River). Further small differences are finally found in the HBV simulations for the Glâne, Jogne and Rotache River basins, the Lake Murtensee and northwestern Lake Lucerne small tributaries, and the Emme River floodplain. The source of these differences is most likely found in the parameter set-up of HBV. Numerical problems (triggered by different hardware or parallelisation) can be ruled out almost certainly because the theoretical formulations behind HBV are not demanding in this respect.

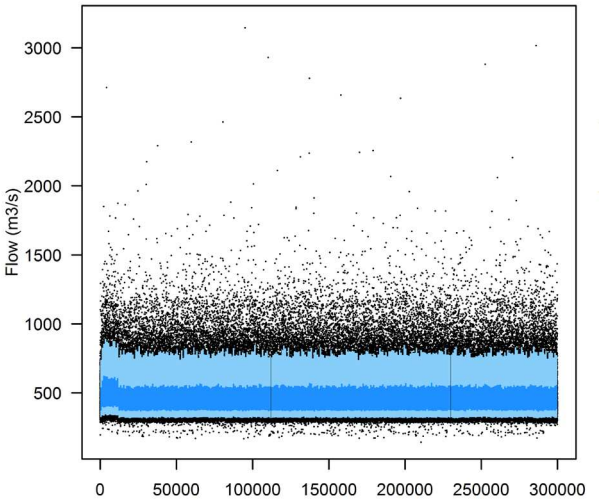


Figure 89 Inconsistency found in January 2019 re-simulations, shown at SSJSSB (Saane River) for the median parameter cluster. Blocks 2–12 are now slightly higher, whereas they were substantially lower before. The plot shows all simulated annual maximum floods (AMFs, black dots), and the 50% and 90% intervals (blue and light blue, respectively) of the AMFs for blocks of 1000 years.

The differences between the simulations of May 2018 and January 2019 are small, and impacts on large floods are mainly found downstream of the Saane River. For assessing these impacts, the 1 000 year flood was estimated from 289 000 AMFs of the 2018 simulation (no simulation error, but slightly different model set-up) and 11 000 AMFs of the 2019 re-simulation. Differences are largest at SSJSSB, where the 2019 simulations show a value 9% higher than that the 2018 simulations. Differences decrease further downstream, with +5% at SSBSSC (Aare River at Golaten), 0% at SSESSF (Aare River at Aarburg) and -2% at SSHSSI (Aare River at Untersiggenthal, Stilli) (Table 5).

Table 5 HQ_{1000} estimated from 2018 simulations (289 000 years unaffected by simulation error) and 2019 re-simulations (11 000 years).

Site	2018 simulations, 289 000 years	2019 simulations, 11 000 years
SSJSSB (Saane River)	1367 m³/s	1496 m³/s
SSBSSC (Aare River at Golaten)	2040 m³/s	2156 m³/s
SSESSF (Aare River at Aarburg)	1625 m³/s	1624 m³/s
SSHSSI (Aare River at Untersiggenthal, Stilli)	3760 m³/s	3679 m³/s

A full run of the remaining 289 000 years in the 2019 simulation environment would have taken an excessive amount of time (an estimated 2–3 months). In view of the small differences between 2018 and 2019 results and the need to move forward, it was decided to discard the affected 11 000 years and to continue with 289 000 years at hand. Also, the events selected from the 2018 simulations were retained and not replaced with a new selection of events from the 2019 simulations. However, the return periods of the events selected from the 2018 simulations were adjusted to correspond to the return periods these events would likely have attained in the 2019 simulations. As a basis for this correction, a reduced amount of simulations was still done with the 2019 set-up, covering the first 100 years of each 1000 year simulation block, for a total of 289 unaffected scenario blocks. Thanks to this, a paired dataset covering 28 900 years with both 2018 and 2019 parameter set simulations was available (see Chapter 4.1).

4.1 Details of adjustment

In the following, the adjustments applied to the 2018 simulations are presented. The general algorithm is provided along with some diagnostic plots.

4.1.1 Data

There are several data series that must be considered in the adjustment:

1. 289,000 years of simulated values from 2018 (preliminary simulations) with median, low, and high parameter sets. The original series had an additional component of 11,000 years, which had to be discarded because a previous version of weather generator inputs was used.
2. 28,900 years of simulations from 2019 (definitive simulations) composed of the first 100 years of each 1000-yr block for the median parameter set only. The “definitive” refers to modified parameters in HBV and RS Minerve. These can be compared to the same floods from the “preliminary” series to look for differences. Only 28,900 years are available because the remaining 1,100 years are in the period using previous weather generator inputs and cannot be paired to a “preliminary” simulation of the same years.

3. 11,000 years of “definitive” simulations that were intended to replace the missing values in the “preliminary” simulation. These are not paired to any valid “old” floods.

4.1.2 Method

There are several possible ways to adjust the floods. The method selected is essentially a moving window average correction, where the window size and location is defined based on the return period of the flood. Because of higher noise for smaller floods, the average includes more points in that range (ca. 4000) than at the very extreme flows (ca. 3). Figure 90 outlines the general algorithm used in the correction.

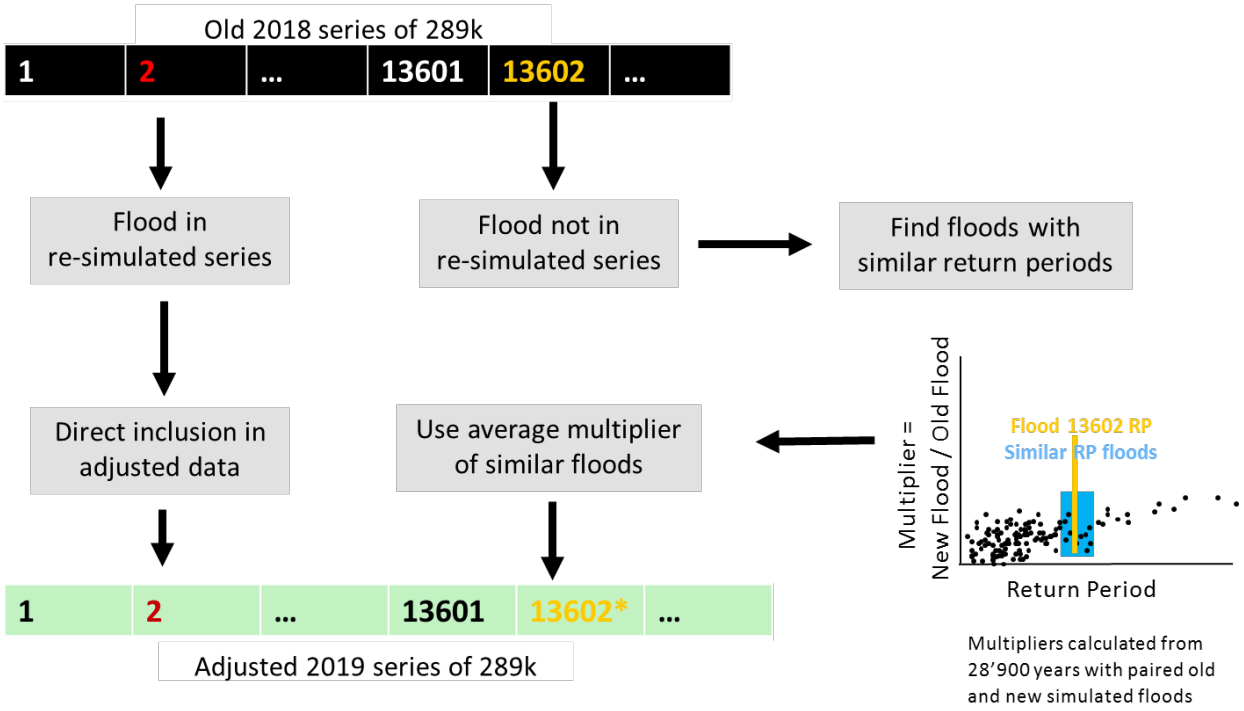


Figure 90 Illustration of the adjustment algorithm. Flood 2 would be in the re-simulated, definitive series. Flood 13602 would not be represented in the re-simulated series, so the average correction is applied.

The first step in the process is to find the multiplicative correction in flow based on the pairs of median parameter set data (dataset 2 of 28 900 years of definitive data and the corresponding floods in dataset 1 of preliminary data). A multiplicative correction was estimated, not an additive one, although over small ranges in the flow the result should be similar. Because it is multiplicative, it would be equivalent to a percentage increase after a small transformation of the computed correction.

Next, the mean value of the multiplier was calculated as a function of the return period of the flood from the preliminary dataset. The number of points used in the correction, n, depended on the return period in years, T, as follows

$$n(T) = 4^{6-\log_{10} T}$$

For a 1-year flood, the number of points is 4096, but for a 10,000-yr flood the number of points is 16. The window is defined as the floods with the closest logarithm of the return period to that being evaluated. This will tend to include more floods with return periods below the nominal return period being adjusted, but due to the shape of the data using the closest points in the flow dimension will have a similar tendency.

Assigning the return periods based on the old dataset essentially stretched them to range from 1 to 289,000 years because there was regularly a “top event” in the re-simulated series of 28,900 years (dataset 2). Several examples of the generated correction functions for transfer points of interest are provided in Figure 91.

The average correction is not guaranteed to be monotonic because it is tracking the average behavior. SSASSB shows some regions of flows where the similar floods increased, and some where they decreased. This can lead to a reordering of the series in terms of peak flow.

For floods that were included in the re-simulated series, the re-simulated flood was given as the re-simulated value and not estimated. This should provide more information than the average correction because the exact value is known.

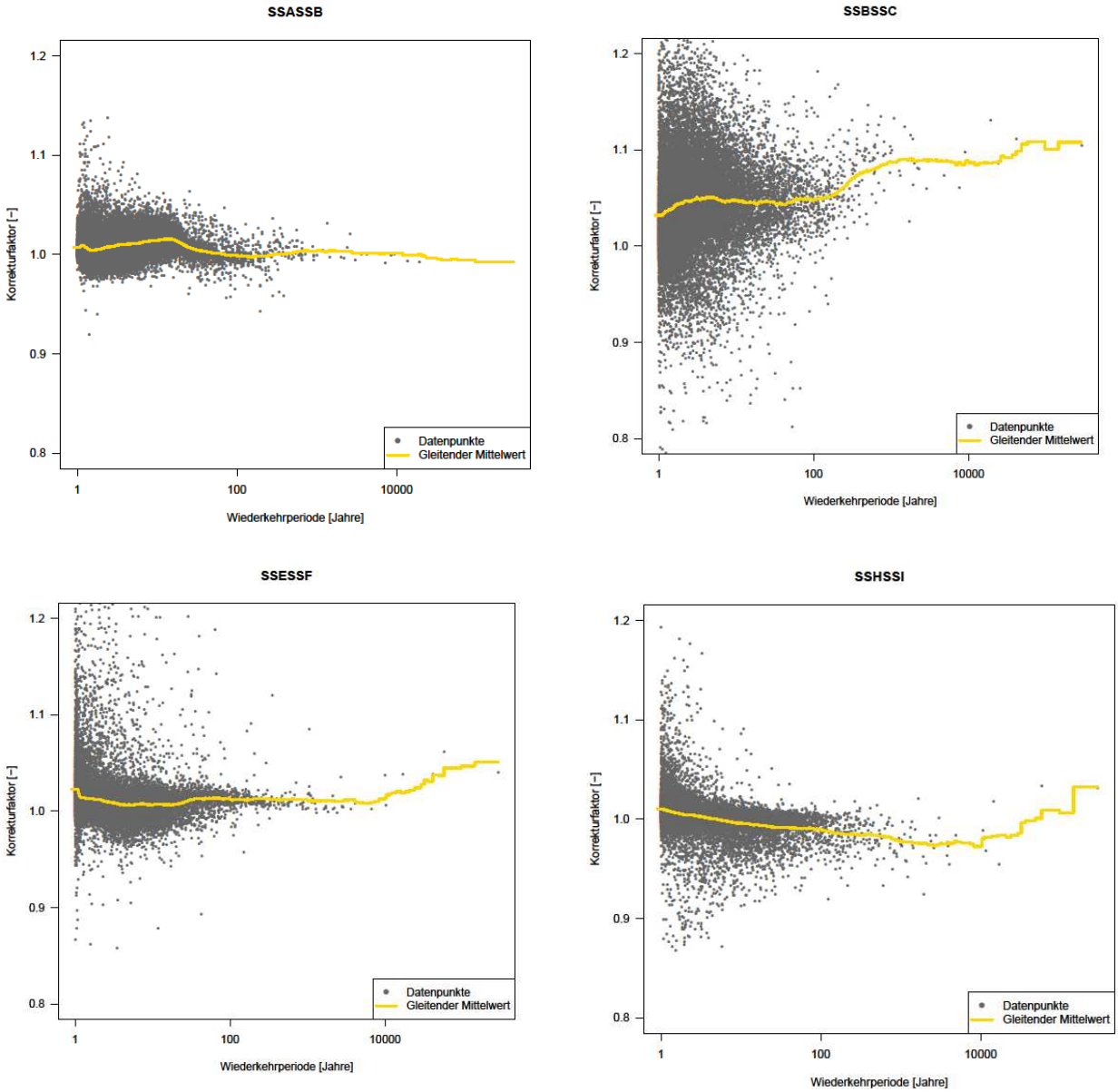


Figure 91 Example correction functions as a function of flow in the preliminary parameter set for SSASSB, SSBSSC, SSESSF, and SSHSSI (Aare Halen, Aare Golaten, Aare Aarburg, Aare Stilli).

Also, there is no correction when the distributions are identical after rounding. This is indicative that there was no change in the parameters that resulted in a difference in the floods, so any changes are presumably negligible. The sites where the distributions are the same to within rounding are LorReu (a small watershed that feeds into the Reuss), SSMSSH (outlet of the Limmat River), and nearly ZueSee (outlet of lake Zurich, only 4 cases with different values). At the other transfer points, the number of cases with the same values between the preliminary and definitive parameter sets within rounding was lowest for SSJSSB (132 cases).

There are two major confluences in the system. At these locations, the flow at the transfer point downstream of the confluence could be infeasible. There is no constraint in the correction algorithm that limits the peak flow downstream to less than or equal to the sum of the upstream peaks. Therefore, this constraint was implemented. The maximum "overestimation" at the Saane-Aare confluence was about 60 m^3/s before the correction, which is not very large given the range of flows considered in EXAR. This does imply that any diffuse inflows between the upstream transfer points and the post-confluence transfer point may be underestimated.

4.1.3 Results

The adjustment algorithm was applied to the median data. The K-S test was used to compare the 11,000 years of definitive data to the original 289,000 years of preliminary data and 289,000 years of adjusted data. The 11,000 years series is independent of the other two series because it was not used to develop the correction and there is no correlation in the simulations over this time scale. The results are summarized in Figure 3, which includes points that were estimated to be the same within rounding.

Generally, the correction increases the similarity of the distributions of the 289,000 years to the 11,000 years of definitive parameter set data, as is shown by most points lying below the diagonal line, which indicates the p-value of the difference in distributions is reduced. Additionally, 11 of the 19 Transfer Points are not statistically different at the 1% level including SSHSSI and SSASSB. SSESSF moved closer to similarity, although it is still not passing the 1% test.

The algorithm does not seem to capture the behavior at the outflow of lakes in many cases (BieSee, VieSee, LacGru). Lake regulation could be one reason for the reduced performance at lakes and possibly the non-monotonicity of transfer points downstream of lakes. The lake regulation might include the flow downstream (e.g., Lake Biel, Lake Thun), which is difficult to account for in this correction.

Note that with this number of observations, 1% statistical significance is equivalent to a difference in the cumulative distribution functions of 0.0147. This does not indicate where in the distribution the maximum difference occurs. It could be for the small floods and there is no difference in the upper tail, or vice versa.

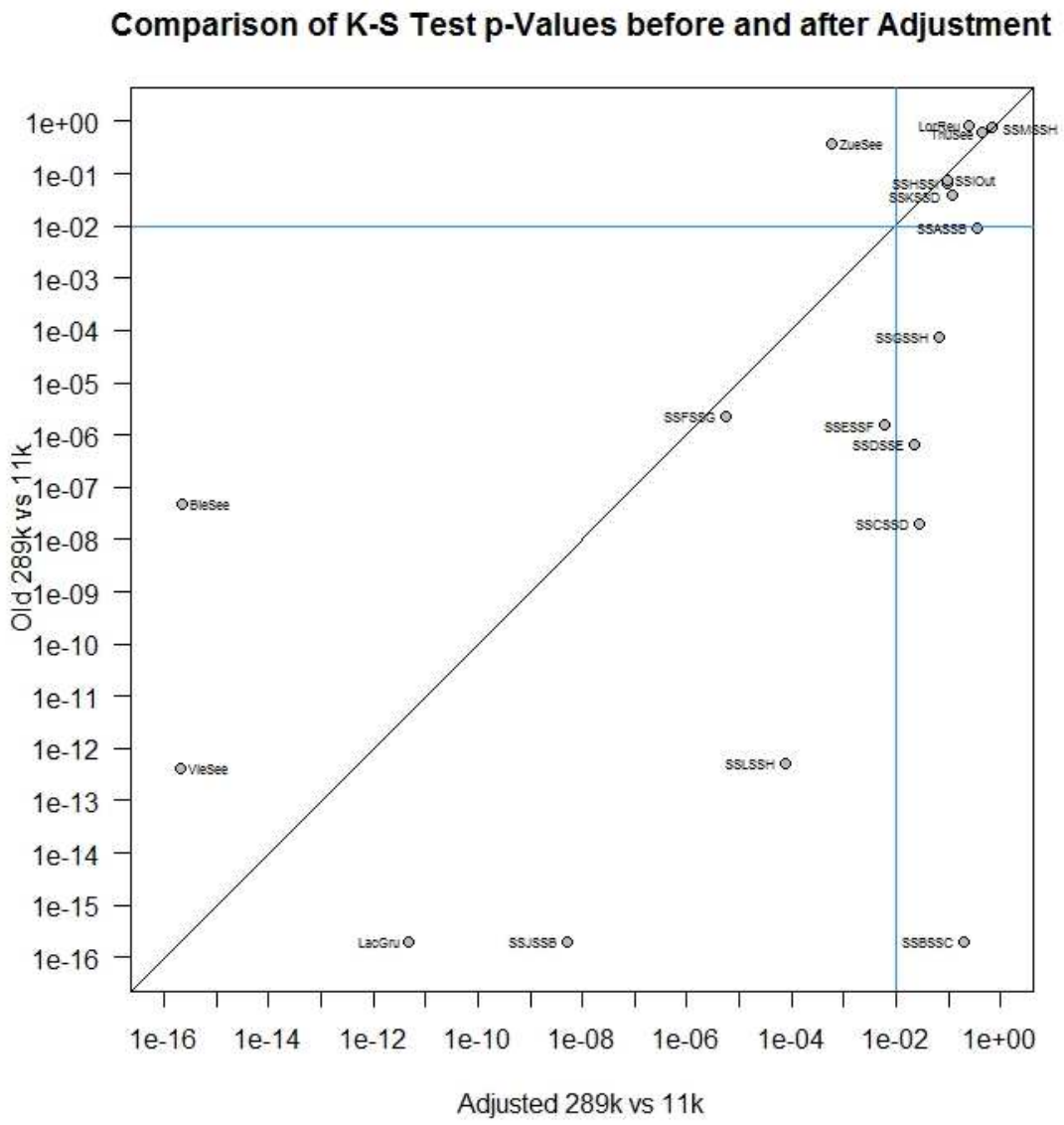


Figure 92 Comparison of K-S tests of similarity of distributions before and after the adjustment. Points below the diagonal line indicate adjustment increased the similarly of the 289,000 year series to the 11,000 years of definite simulations. The blue line represents the threshold for 1% significance.

One example of where the adjustment method does not work well is VieSee (outlet of Lake Lucerne). This location shows some unusual behavior around 250 m³/s, which was possibly translated into the cumulative distribution function. In this case, the adjustment should be applicable to the high floods because there is no correction there, but there would be questions of applicability for events less than the 10-yr flood.

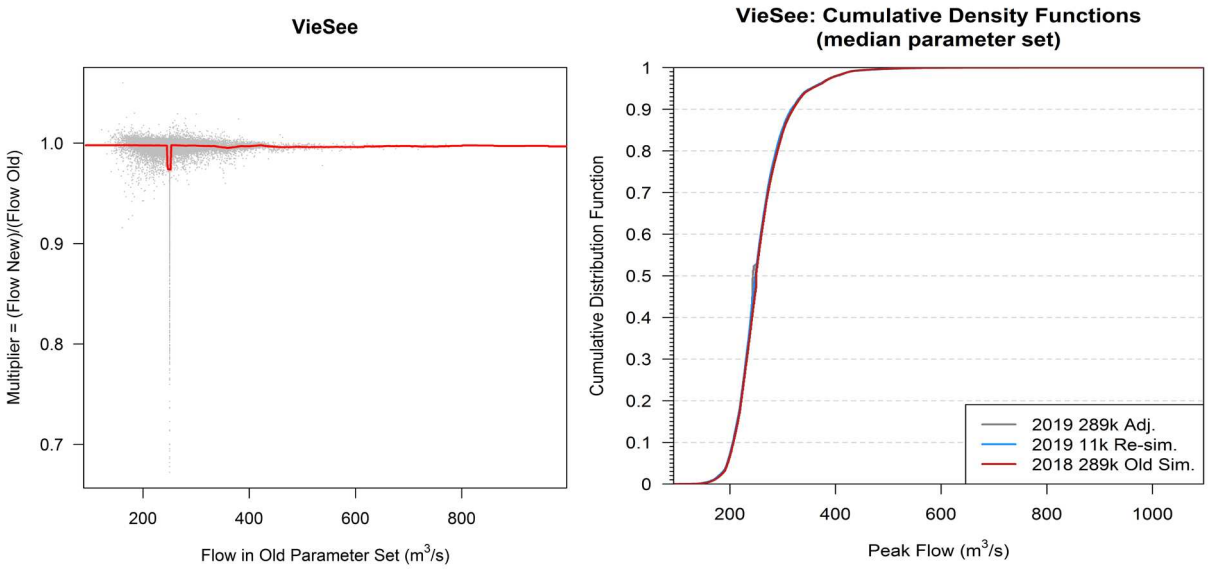


Figure 93 Example of multiplier correction function and cumulative distribution function for VieSee (outflow of Lake Lucerne). The nearly vertical line at around 250 m³/s on the left plot could correspond to the "kink" at around 250 m³/s on the right plot, the likely location of the maximum difference.

4.1.4 Adjustments to Low and High Parameters

There are two other hydrologic parameter sets considered in EXAR. The algorithm proceeded as previously, except the average correction as a function of the return period was used throughout. There were no simulations of the same flood events with a definitive low or high parameter set, so a specific adjustment to those floods could not be estimated.

4.1.5 Impact of Reference Floods and Frequencies

Changes to the floods will naturally have an impact on the frequencies of given floods. Tables 1 and 2 summarize the impact of the adjustment on frequencies of simulated flows (Table 1) and the changes to flows that would have the same exceedance frequency (Table 2). Generally, the flows are increasing with the largest changes at SSJSSB and SSBSSC. SSHSSI shows slight decreases in the extreme, largely because the flow from the Reuss decreased for extremes. A summary for all TPs can be found in Table 9.

Table 6 Frequency increases as a factor of the nominal floods based on the 2018 preliminary series. Values greater than 1 indicate increased exceedance frequency. Values are based on the median parameter set only. For example, 1.1 indicates 10% relative increase in the nominal frequency (1E-3/a before, 1.1E-3/a after)

TP	HQ1k	HQ10k	HQ100k
SSASSB	1.1	1.0	0.5
SSJSSB	2.3	2.6	3.5
SSBSSC	2.0	2.3	3.0
SSESSF	1.1	1.2	1.5
SSHSSI	0.7	0.8	1.0

Table 7 Percentage increases of flows based on the 2019 adjusted data compared to the 2018 preliminary series. A value of 10 would indicate the reference 2018 flow value would increase by 10% to have the same exceedance frequency as in the 2019 series (new frequency = 1.1*old frequency).

TP	1E-3/a	1E-4/a	1E-5/a
SSASSB	0.4	0.1	-0.8
SSJSSB	13.1	11.7	12.9
SSBSSC	8.8	8.9	10.2
SSESSF	1.2	1.3	4.7
SSHSSI	-2.4	-2.6	0.9

5 Overflow of Rhine River into Aare River at Mels

Daniel Viviroli

A critical question for the entire project is whether the Rhine River, in an extreme scenario, can overflow towards the Seez River and Lake Walensee in the region of Mels. If this was possible, significant amounts of additional flow could enter the Aare River system and thus change the assessments made in this report. As a basis for answering this question, elevation profiles were extracted from the 2 m resolution land surface model SwissALTI3D. They were drawn following the direction of a possible overflow, with a spacing of 50 m (see Figure 94).

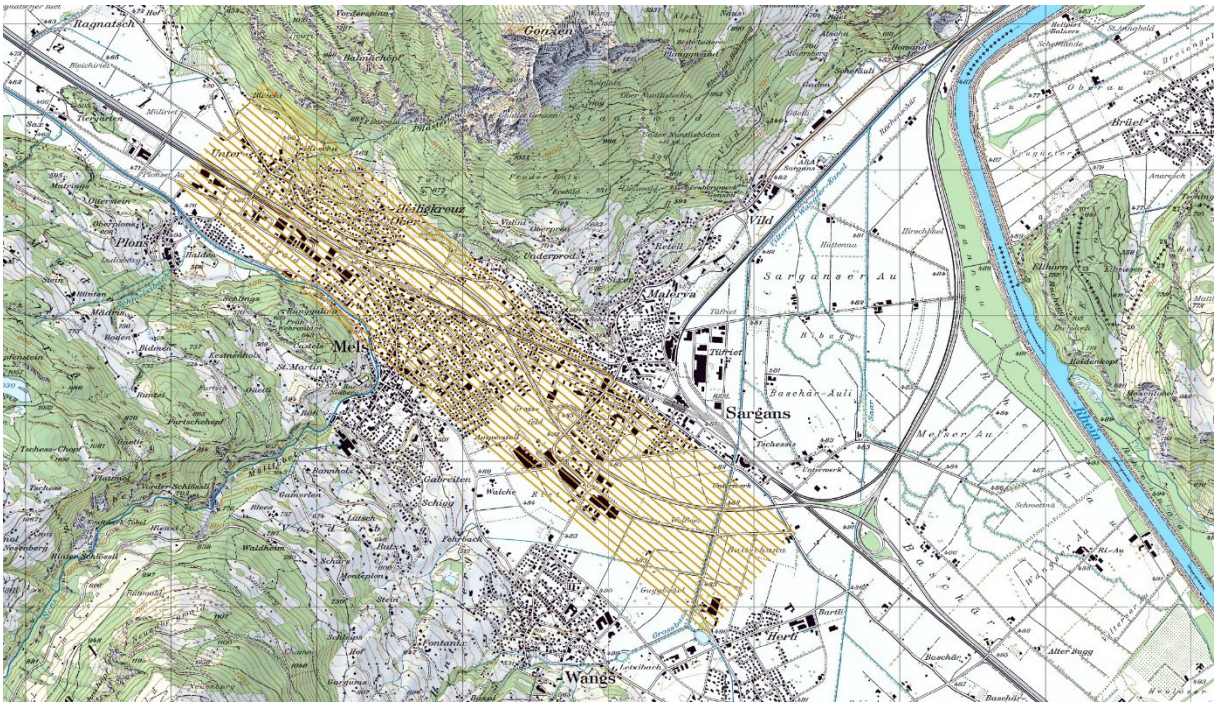


Figure 94 Location of profiles shown with orange lines.

Results show that the rise separating the Rhine and Aare River basins has an effective elevation difference of roughly 6 meters (minimum elevation on Rhine River side: ~480.5 m a.s.l., maximum elevation around Mels: ~486.5 m a.s.l.; see Figure 95). Given the morphological situation of the Rhine River valley floor, the only kind of flood that could possibly surmount such an elevation difference would be a massive flood wave. This would require multiple dam breaks further upstream in the Rhine River, e.g. due to an earthquake. Even in this unlikely case, it can be assumed that additional flow is

rather splashing over towards the Seez River, rather than continuously overflowing with important volumes of water. Lakes Walensee and Zurich would buffer this spill over strongly before they reach the Aare River. It can thus be concluded that an overflow is not relevant within EXAR.

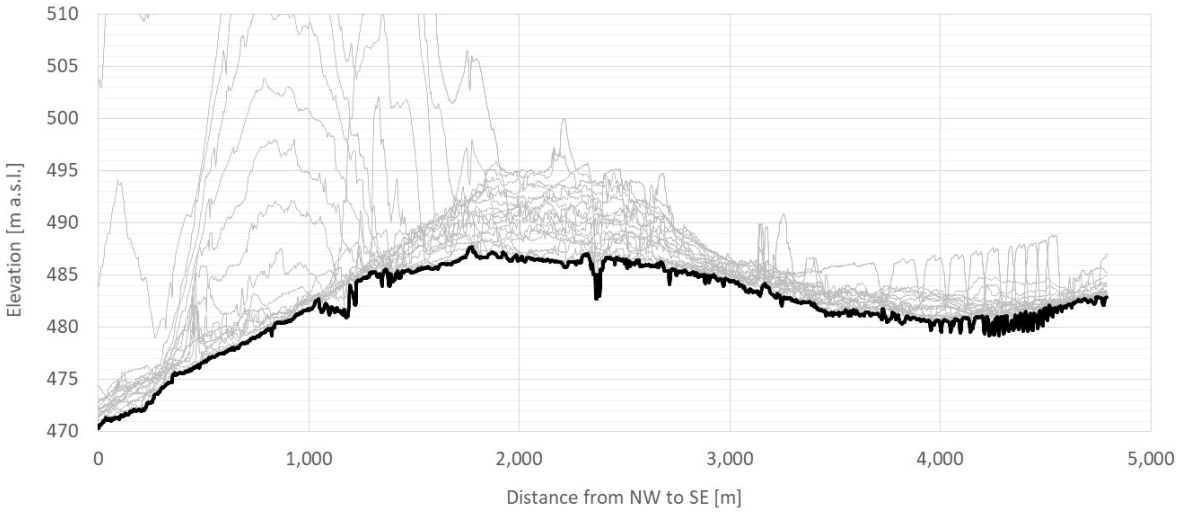


Figure 95 Elevation of profiles shown in Figure 94, cutting from the Northwest (Aare River basin, left) to the Southeast (Rhine River basin, right). The grey lines refer to individual profiles, the black line represents the lowest point of all profiles.

6 Discharge uncertainty (rating curve)

Ida Westerberg

Measurements at the gauges are streamflow levels rather than discharge volumes. Hence to use them for hydrological modeling in calibration and validation the levels are first transformed into volumes by means of rating curves. These rating curve bear uncertainties, which were estimated in this chapter for several streamflow gauges. The analyses give an idea of how much the uncertainty of the rating curve can affect the results for the flood events that are generated in the project.

A general description of the methods and an overview of the results is given below in Chapter 6.1 using the Orbe River as an example. Detailed results for the further gauges examined follow in Chapters 6.2-6.17.

6.1 Methods and overall results, with example of Orbe River

The rating curve and gauging data for the BAFU stations for the period 1980–2014 were analysed with respect to temporal variability, number of gaugings, amount of extrapolation, bankfull levels, etc. At all analysed stations there were a large number of rating curves that had been used historically, typically around 40–60 rating curves per station. Since it was not feasible to estimate uncertainty for so many curves individually, they were for each station grouped into time periods that could be considered to have stable stage-discharge relationships (with particular focus on the stability at high and medium flows). Some stations and time periods for which there was large temporal variability, or not enough gaugings for reliable uncertainty estimation, were excluded from the uncertainty analysis. In total the estimates were made for 16 of the BAFU stations. For 11 of the stations a single rating curve was used, two periods were used for three of the stations, and three periods for two of the stations. Discharge uncertainty was estimated for the period 1980–2014, or if there was temporal

variability in the earlier period only for 1990–2014, which was the time period for which the data are used in the calibration of the HBV hydrological model.

Manning estimates of high flows for the highest stage were made to constrain high-flow uncertainty at stations with large rating curve extrapolations, where sufficient cross-section information was available. The uncertainty in the Manning discharge values was estimated by assuming an uncertainty of +/- 4% for the discharge measurement, and +/- 2 cm for the stage measurement. These uncertainty distributions were normally distributed and were incorporated into the estimation as an additional flow gauging for the highest stage (with the estimated discharge uncertainty).

Rating curve uncertainty was estimated in a Monte Carlo uncertainty analysis using the Voting Point likelihood method (VPM; McMillan and Westerberg, 2015; Westerberg et al., 2016). The method essentially consists of estimating multiple feasible rating curve parameter sets that are compatible with the gauging data. The measurement uncertainty in stage and discharge gaugings was taken into account in the estimation. These error magnitudes were set after discussion with BAFU: using a uniform error of ±5mm for stage for low to medium flows and ±20mm for high flows. For discharge we used a normally distributed error with 95% bounds at ±4% for current meters, ±15% for float gaugings, ±6% for ADCP, salt dilution and other techniques. Many of the official rating curves were polynomial curves, but to be able to apply the VPM method it was necessary to estimate power-law rating curves. Single or multi-section initial rating curves were therefore estimated from the gauging data, using information about river cross-sections and transition stages for the latest official rating curves to set the section breakpoints. These initial rating curves were then used to initialise the estimation of the uncertainty in the rating curve parameters with the VPM method. In total we estimated 40 000 different rating curves.

The results show that in general the uncertainty is low in the gauged range with well constrained distributions and the uncertainty increases for high flows (Figure 96, results for all stations are shown in the following subchapters). For low flows the absolute uncertainty values are small, but the relative uncertainty (i.e. percent uncertainty) is higher at stations with more spread in the low-flow gauging data. There is greater uncertainty at high flows for the stations where there is spread in the high flow gauging data and those where there is larger extrapolation. The uncertainty in the extrapolated part is constrained by the rating-curve parameter priors in combination with the constraints provided by the gaugings at lower stages. For the stations were Manning estimates were made, these constrained the uncertainty in the extrapolated curve further. The parameter priors were set to standardised ranges, similarly to Westerberg et al. (2016), and adjusted for each station if necessary. The propagation of rating curve uncertainty to the hourly discharge time series is illustrated for the large flow event in January 2004 at the Orbe station (Figure 96, right). The figure shows how the discharge uncertainty is well constrained for low and medium flows and increases in the extrapolated part above the maximum gauged flow (note that the rating curve shows the instantaneous discharge and the time series the hourly discharge).

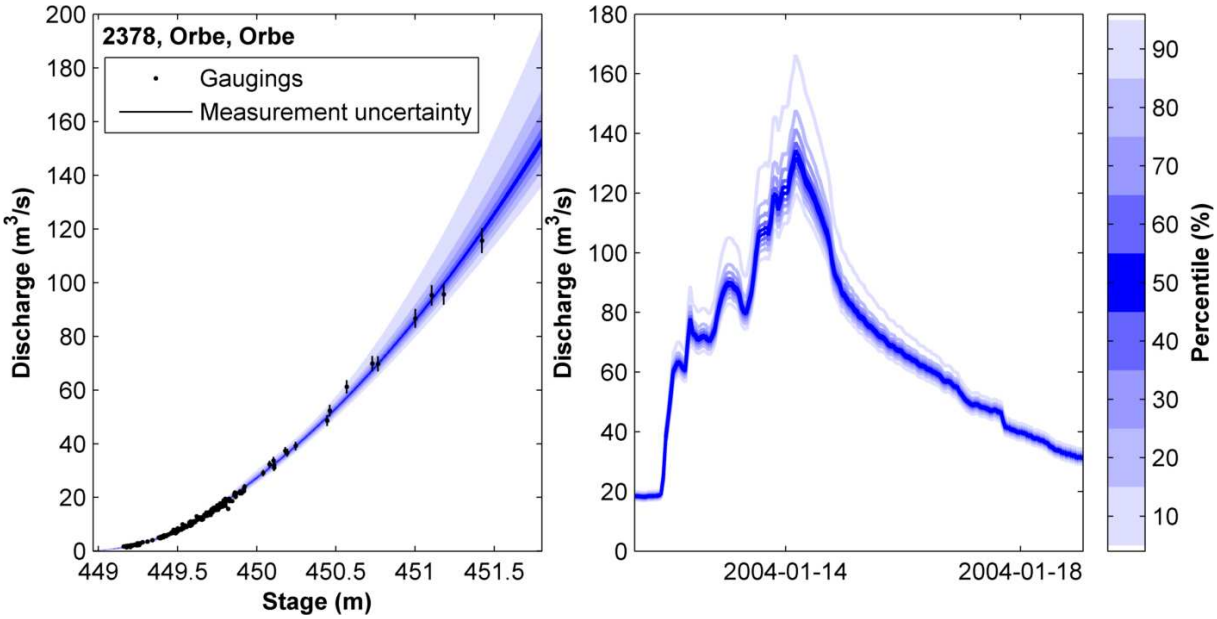


Figure 96 Rating curve with estimated uncertainty for the Orbe station on the river Orbe (left) and the propagated uncertainty in the hourly time series for the maximum flow event (right). The black dots show the gauging data and the black lines the 95% bounds of the discharge measurement uncertainty. The blue bands show the uncertainty distribution of discharge from all of the 40 000 estimated rating curves at each stage value, from the 5th to 95th percentiles. Narrower blue bands means that there is a higher probability density for the discharge.

The average uncertainty for the maximum hourly flow was ±24% (half-width of the 95% uncertainty bound, i.e. the 2.5–97.5th percentile interval), varying in the range ±13–40% between the stations (Table 8). Several stations had heavy-tailed uncertainty distributions for the maximum flow with an up to 7% smaller 90% uncertainty bound. The uncertainty was highest for stations and rating-curve periods with large extrapolations where it had not been possible to make Manning estimates. The uncertainty in high-flow signature values generally increased with flow and was on average ±15% for Q1 (i.e. the hourly flow exceeded 1% of the time), ±16% for Q0.1 and ±21% for Q0.01.

Table 8 Uncertainty in hourly maximum flow and high-flow signatures, where e.g. Q1 is the hourly flow exceeded 1% of the time. The uncertainty is calculated as half-widths of the uncertainty interval for the 95% bound (2.5th– 97.5th percentiles) or the 90% bound (5th– 95th percentiles). The maximum flow uncertainty is calculated for the whole time period using all the 40 000 realisations, the signature uncertainties are calculated for 1000 realisations for the period 1990–2014

River and station	Maximum flow 95% uncertainty bound (%)	Maximum flow 90% uncertainty bound (%)	Q1, 95% uncertainty (%)	Q0.1, 95% uncertainty (%)	Q0.01, 95% uncertainty (%)
Broye-Payerne	19	13	14	10	15
Sarner Aa-Sarnen	13	11	12	7	8
Linth-Weesen	23	16	14	24	24
Gürbe-Belp	40	33	22	30	36
Sarine-Broc	40	36	10	13	22
Sihl-Zürich	22	19	17	17	23
Sense-Thörishaus	17	16	13	11	12
Mentue-Yvonand	21	17	14	9	15
Orbe-Orbe	24	19	11	17	24
Kleine Emme-Littau	20	19	16	13	17
Dünnern-Olten	13	11	12	5	9
Emme, Burgdorf	26	21	23	21	22
Wigger-Zofingen	24	22	17	16	22
Kander-Hondrich	33	26	17	26	32

6.2 Broye, Payerne

Discharge uncertainty was estimated for 1980–2014. All the gauging data 1978–2016 were used for this station as there was no conclusive evidence about temporal change. A two-section power-law curve was fitted to the data using information about the transition stage interval. The estimated rating curve uncertainty was in general small, with higher values in the middle of the stage interval where there was a larger spread in the gaugings (Figure 97). The estimated uncertainty aligned well with the variability in the historical official rating curves.

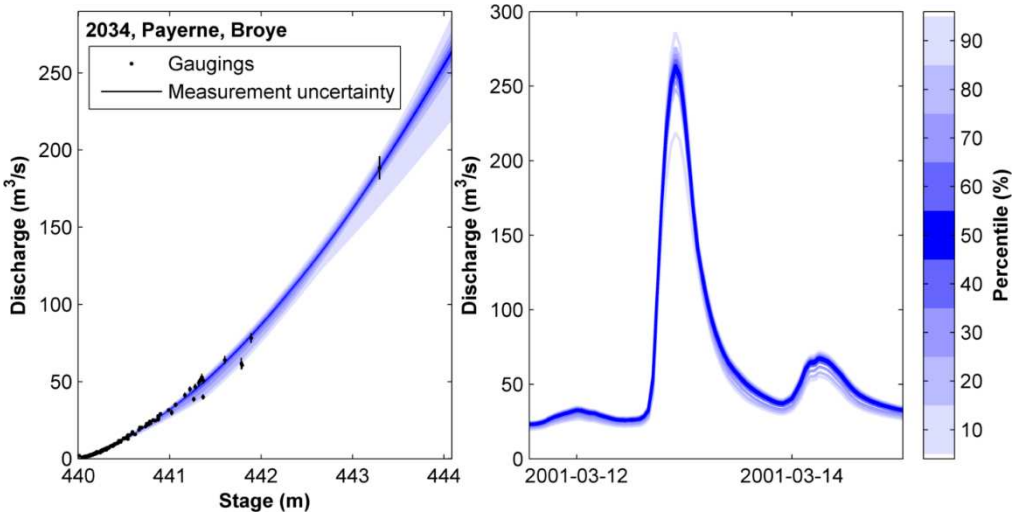


Figure 97 Discharge uncertainty Broye, Payerne.

6.3 Sarner Aa, Sarnen

Discharge uncertainty was estimated for 1980–2014 using all the gauging data for 1974–2016. A two-section power-law curve was fitted to the data using information about the transition stage interval. The estimation was not made for water levels higher than the out-of-bank stage of 470.5 m. An additional flow that bypassed the station was added to the recession of the 2005 flood event, similarly to the official discharge time series.

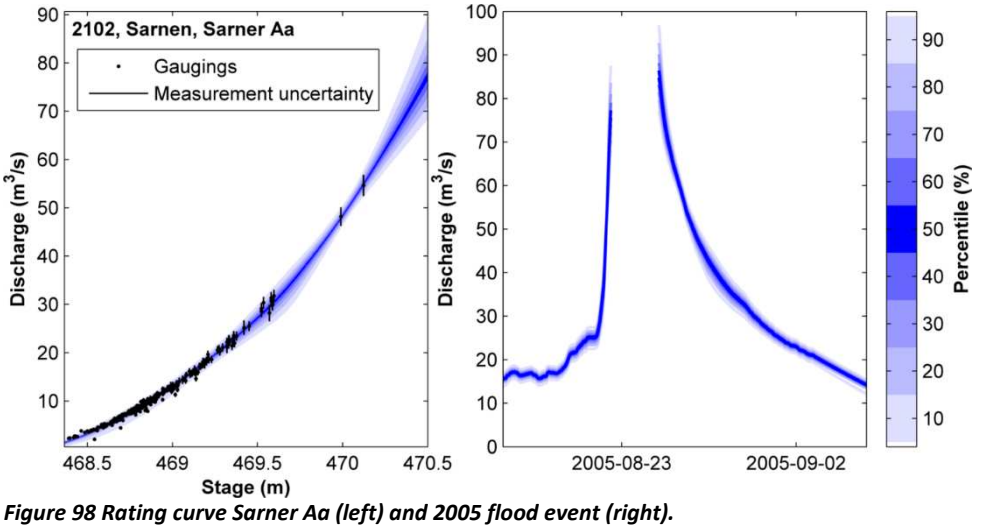


Figure 98 Rating curve Sarner Aa (left) and 2005 flood event (right).

6.4 Linth, Weesen

Discharge uncertainty was estimated for 1980–2014. All the gauging data 1979–2016 were used for this station. A one-section power-law curve was fitted to the data, the estimation was divided into two periods to reflect the change in the stage-discharge relation around 26 April 1999.

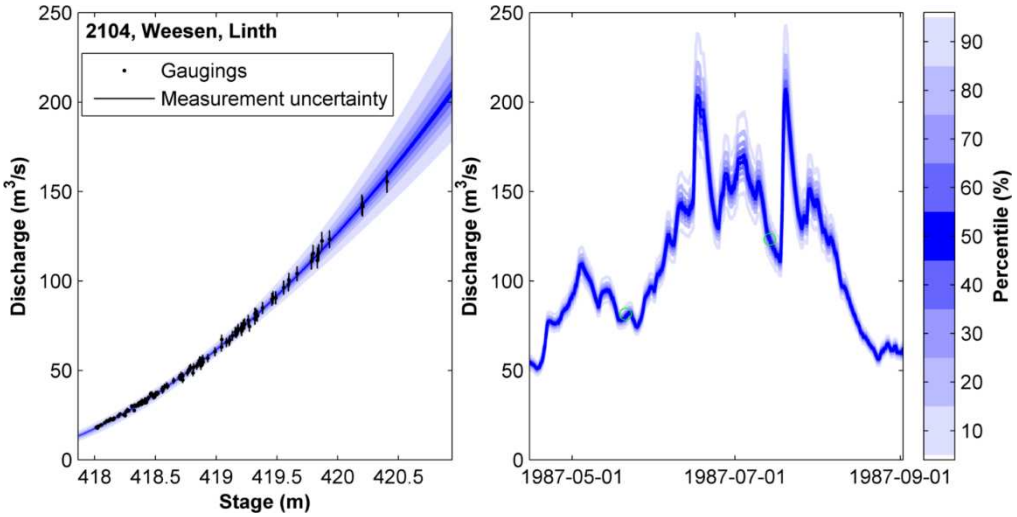


Figure 99 Rating curve Weesen, Linth Aa (left) and results for the first period before 26 April 1999 (right).

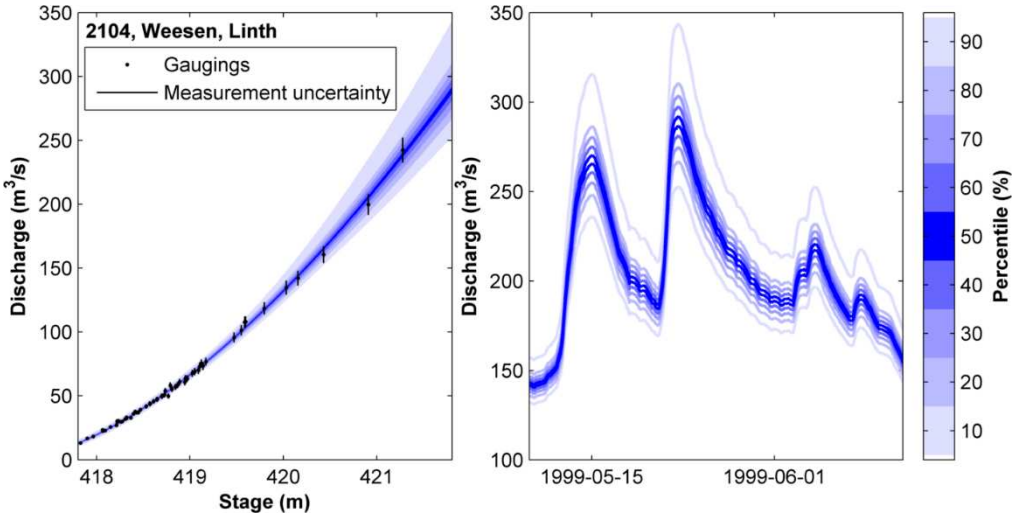


Figure 100 Rating curve Weesen Linth (left) and period for the second period after 26 April 1999 (right).

6.5 Gürbe, Belp

Discharge uncertainty was estimated for 1991–2014. All the gauging data 1991–2016 were used for this station. The data had different stage base levels but it was possible to adjust them to the same stage scale so that they followed the same rating relation. Gaugings in the early part of the period that were recorded in an older stage scale were adjusted to the new scale using the new stage time series data. A one section power-law curve was fitted to the data. Time series uncertainty results are shown for the highest event in the official discharge time series (Figure 101).

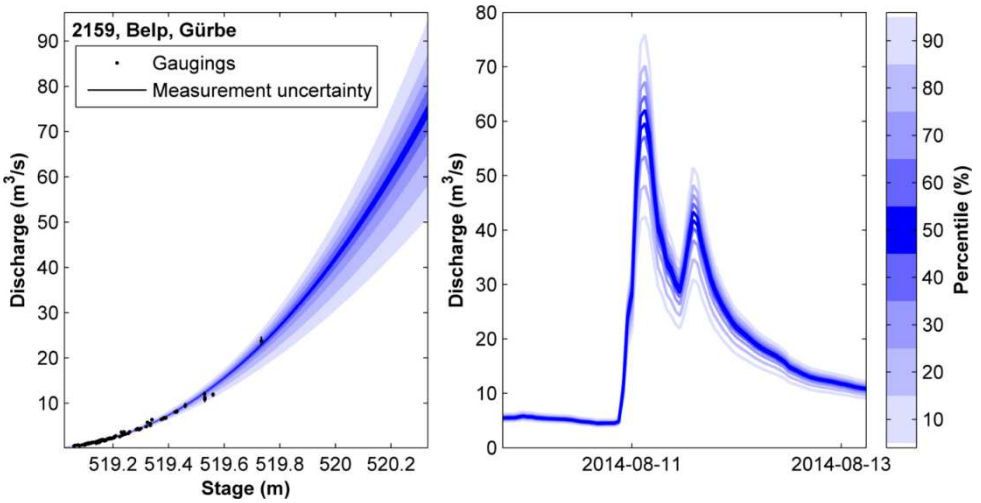


Figure 101 Rating curve Gürbe, Belp (left) and the highest event in the official discharge time series (right).

6.6 Sarine, Broc

There was a large change in the rating in March 1995, and the years before and after 1995 there were differences between the stage time series and the gauging stage values. Therefore, the period 1993–2004 was excluded from the uncertainty estimations. Three different rating curve periods were used: 1980-01-01–1993-10-08, 2004-01-11–2004-10-20, and 2004-10-21–2016-06-06. For the last two periods a Manning estimate was used to constrain the high flows. The estimates were only made up to bank-full stage of 683.8 m, for the earlier period the estimates were only made up to 682.8 m as the rating base level had shifted one meter in-between the periods and the bank-full level of the earlier period was not known. The rating curve for the latest period is shown in Figure 102 together with the discharge time series for the 2007 July flood event.

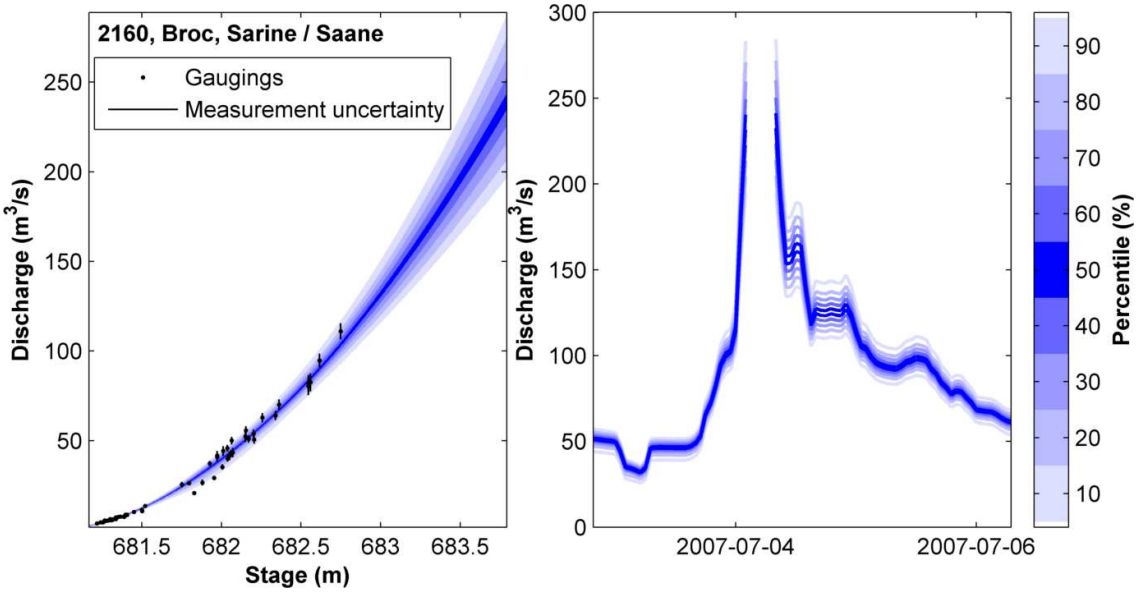


Figure 102 Rating curve Sarine (2004-10-21–2016-06-06) and the July 2007 flood event.

The rating curves for the periods 1980-01-01–1993-10-08 and 2004-01-11–2004-10-20 are shown in Figure 103. For the 2004 curve the Manning estimate constrained the high flow uncertainty.

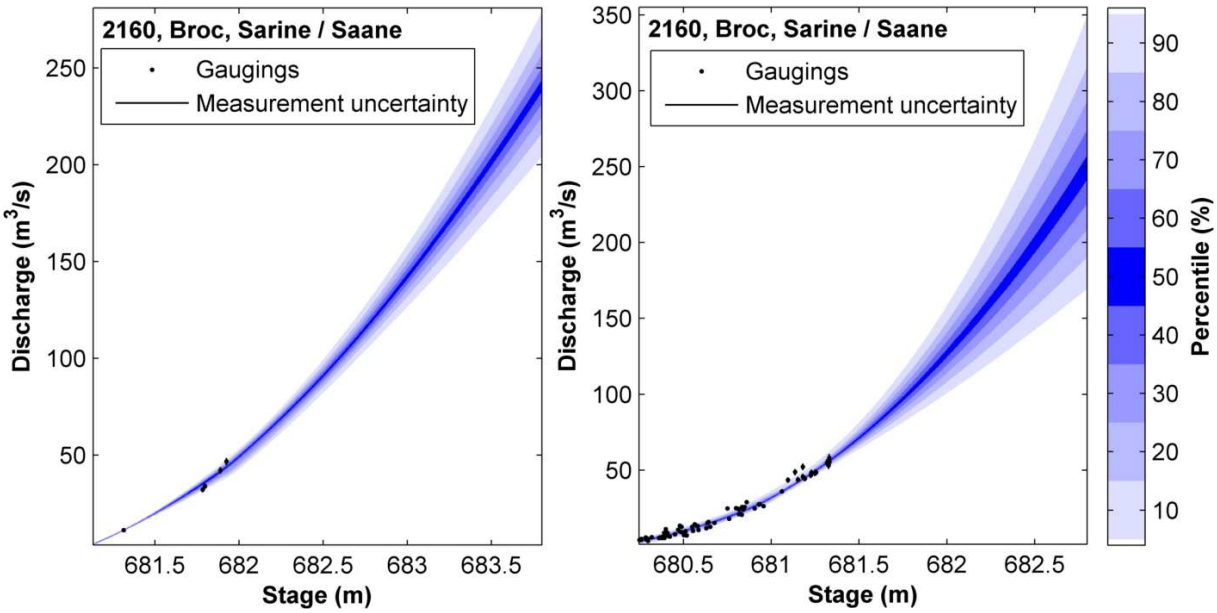


Figure 103 Rating curves Sarine for the periods 1980-01-01–1993-10-08 (right) and 2004-01-11–2004-10-20 (left).

6.7 Sihl, Zürich

Discharge uncertainty was estimated for 1981–2014. All the gauging data 1981–2016 were used for this station (besides two low-flow outliers that fitted the earlier stage-discharge relation). Additional cantonal float gauging was used to constrain the high flows with an estimated uncertainty of ±15% of each gauging. A one section power-law curve was fitted to the data.

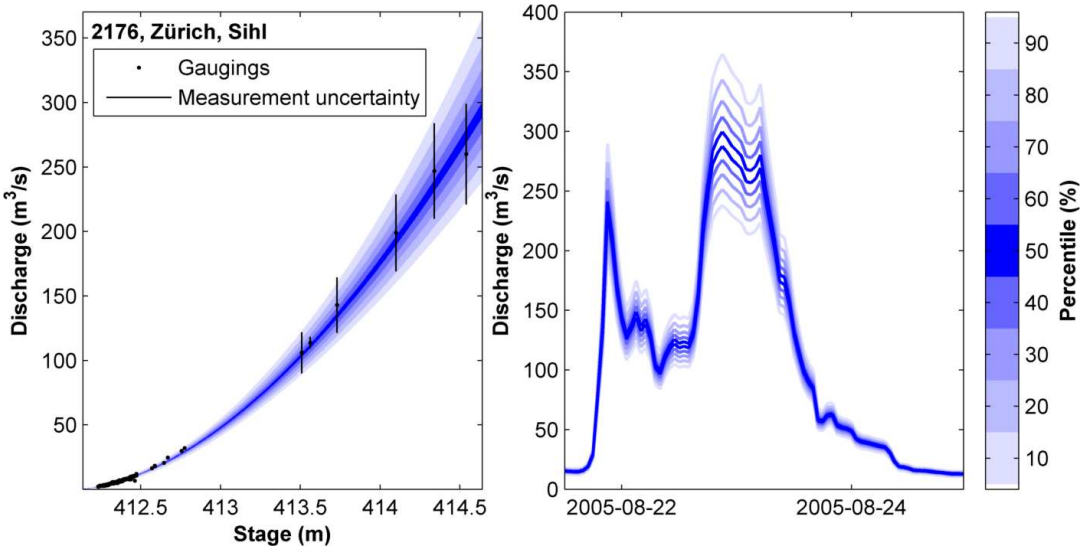


Figure 104 Rating curve and August 2005 flood event for the Sihl at Zurich.

6.8 Sense, Thörishaus

All the gauging data 1979–2016 were used for this station, but the data from 2003-06-04 to 2007-08-08 that had more variability in the gauging data than the official rating curves. Discharge uncertainty was estimated from 1980 to 2014. A three section power-law curve was fitted to the data. A Manning estimate was used to constrain the rating curve in the upper part. The rating curve and the highest flow event in the time series are shown in Figure 105.

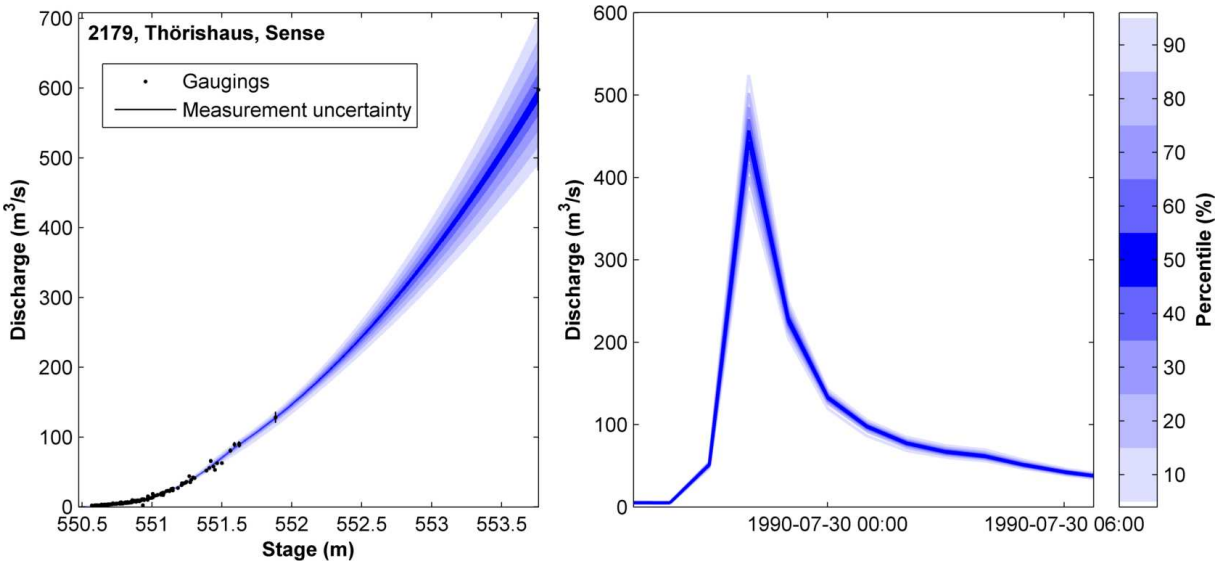


Figure 105 Rating curve for the Sense at Thörishaus (left) and July 1990 flood event (right).

Note that the highest hourly average value (i.e. in the time series) is significantly lower than the highest instantaneous discharge (seen in the rating curve).

6.9 Mentue, Yvonand, La Mauguettaz

Discharge uncertainty was estimated for the period 1980–2014. All the gauging data 1974–2016 were used for this station. A two-section power-law curve was fitted to the data using information about the cross-section to fit the breakpoint. Two periods with suspect flat-line data (2012-02-02–2012-02-24 and 2011-10-26–2011-11-09) were excluded. The rating curve results and the discharge time series for the highest event in the period 1990–2014 are shown in Figure 106 (an even higher event occurred in the 1980s).

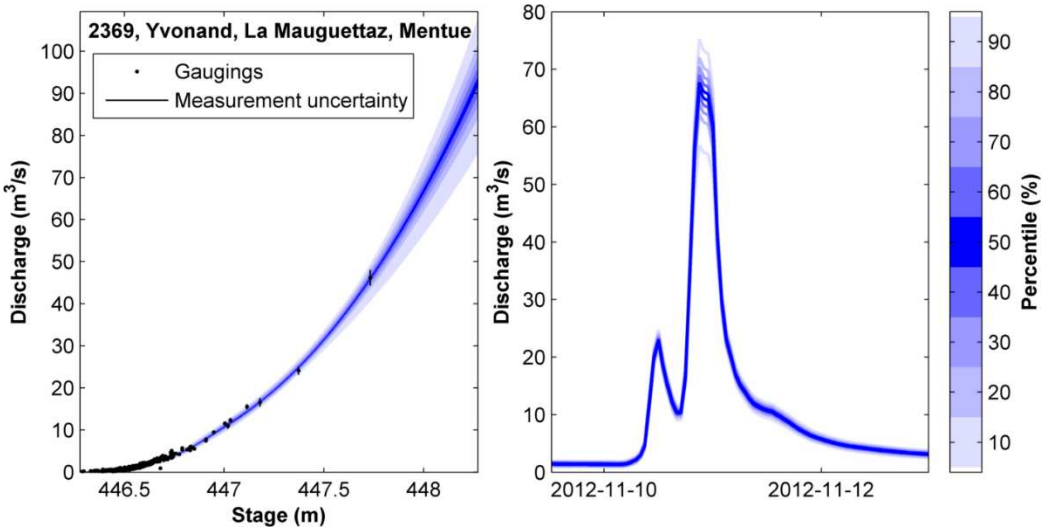


Figure 106 Rating curve for Mentue, la Mauguettaz at Yvonand (right) and flood event of November 2012.

6.10 Orbe, Orbe

Discharge uncertainty was estimated for 1980–2014. All the gauging data 1979–2016 (except one large outlier) were used for this station as there was no evidence about temporal change at the station. A two-section power-law curve was fitted to the data using information about the cross-section from

the Swiss terrain model to set the breakpoint. Results for the rating curve uncertainty and for the highest event in the official hourly time series are shown in Figure 107.

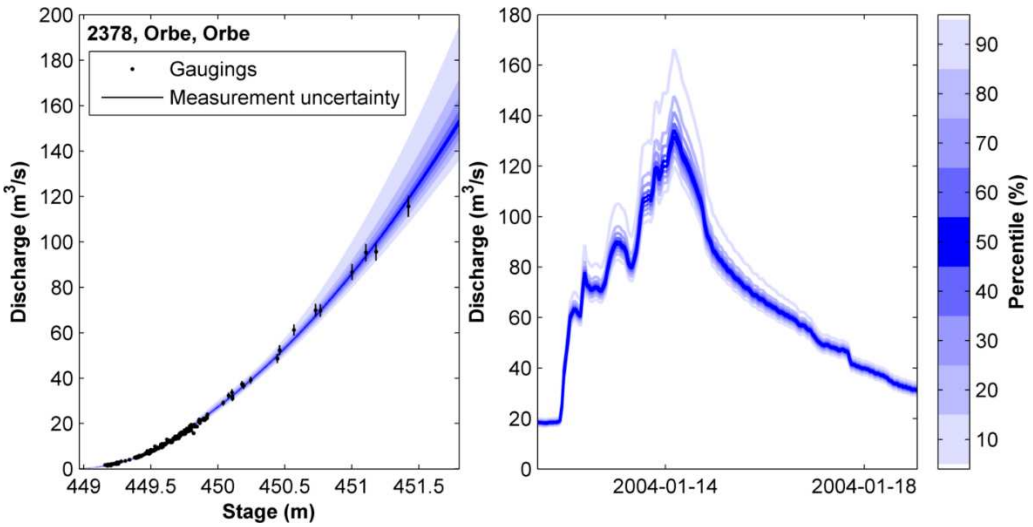


Figure 107 *Rating curve at Orbe, Orbe (left) and flood event of January 2004 (right).*

6.11 Kleine Emme, Littau, Reussbühl

Discharge uncertainty was estimated for 1980–2012. All the gauging data March 1979–December 2012 were used for this station, and the data were divided into two periods as there was a different rating curve during the period 1999-01-15–1999-04-09. Discharge was not calculated for the transition period 9 April to 17 April, when there was a difference to the official discharge data. A two-section power-law curve was fitted to the data using information about the transition stages. The estimations were only made up to bank full stage of 434.3 m. A Manning estimate was used to constrain the rating curve in the upper part. Results for the main rating curve period, with the highest event in the official hourly time series are shown below.

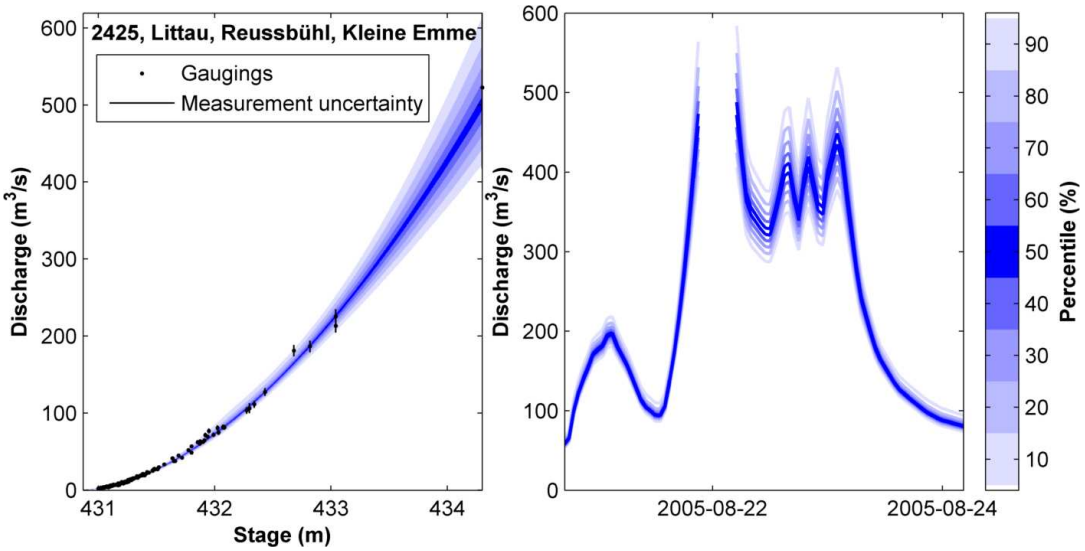


Figure 108 *Rating curve at Kleine Emme, Littau (left) and flood event of August 2005 (right).*

The results for the period 1999-01-15–1999-04-09 with the highest event in the hourly time series are shown in Figure 109.

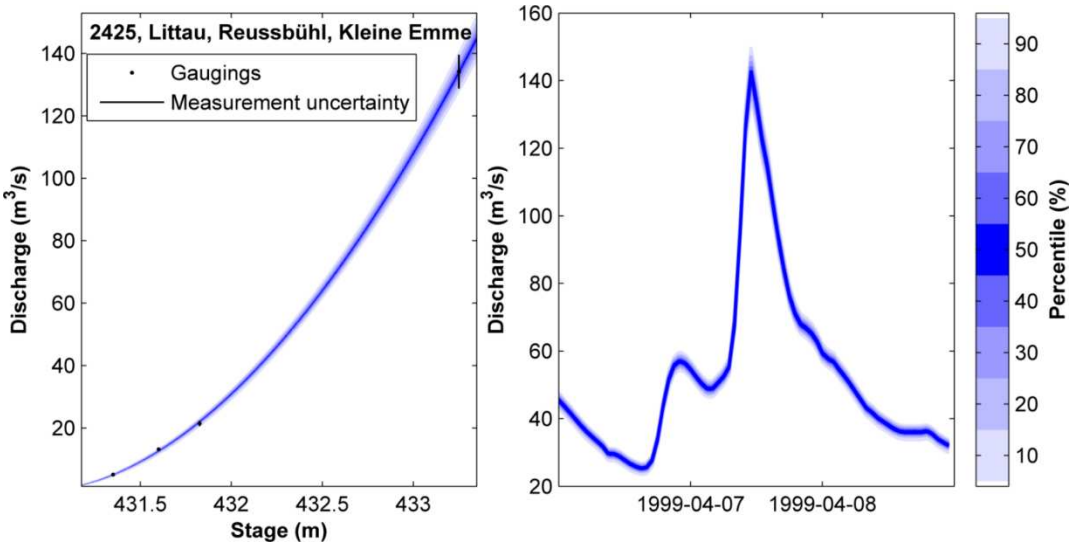


Figure 109 Rating curve for the period 1999-01-15–1999-04-09 at Kleine Emme, Littau (left) and flood event of August 2005 (right).

6.12 Dünnern, Olten

Discharge uncertainty was estimated for 1990–2014, using the gauging data from February 1990–May 2015. A two-section power-law curve was fitted to the data using information about the transition stages.

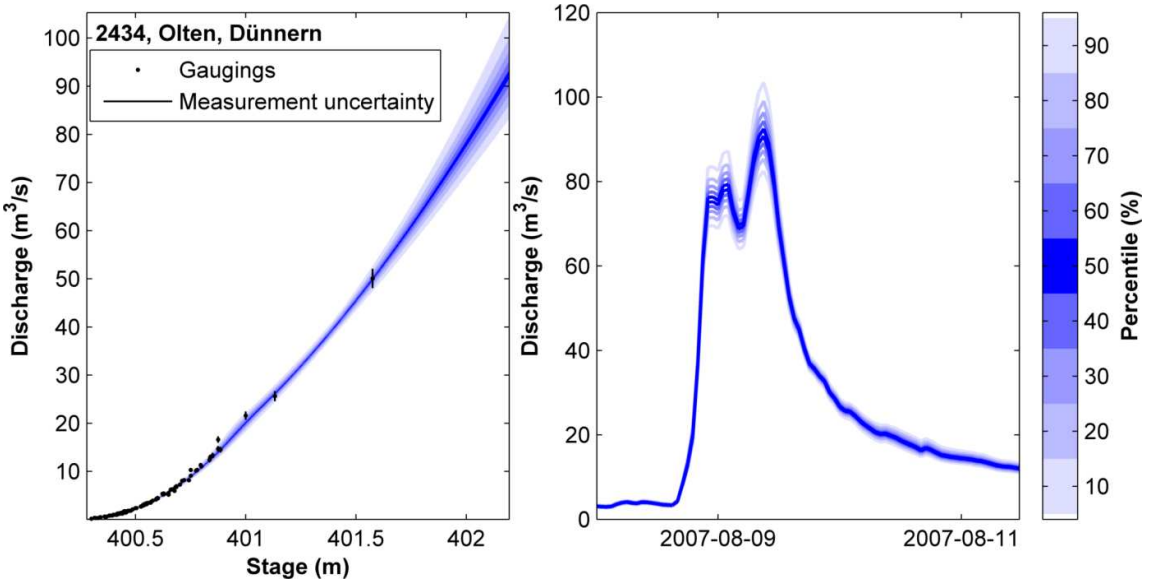


Figure 110 Rating curve for Dünnern, Olten (left) and flood event of August 2007 (right).

6.13 Emme, Burgdorf

Discharge uncertainty was estimated for 1980–1990, using the gauging data for the same period. The station was discontinued at the end of 1990. A Manning estimate was used to constrain the rating curve in the upper part. A one-section power-law curve was fitted to the data and the stage base level was adjusted so that the whole period had the same base level.

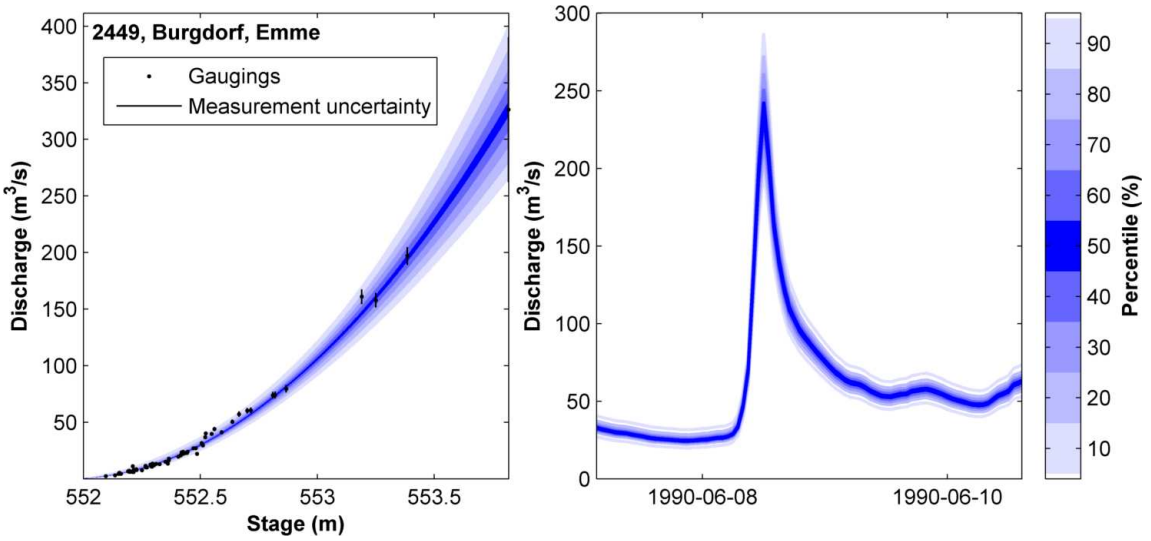


Figure 111 *Rating curve for Emme, Burgdorf (left) and flood event of August 2005 (right).*

6.14 Wigger, Zofingen

All the gauging data for the period 1979–2013 were used as there was no conclusive evidence about temporal change at the station; there were some spread in the high flow gaugings, however, all the last three high flow gaugings come from the same flow peak and a previous hydraulic hazard study with HEC-RAS (Schälchli, Abegg + Hunzinger AG, 2007) reports a higher rating curve than the current official one. The discharge time series is affected by short-term river regulation and these periods were manually removed from the water level record before calculating the discharge.

There was an extrapolation of the rating curve of over 1 m during the analysed period, therefore Manning estimates of the discharge for the maximum stage was made and included in the rating curve uncertainty estimation.

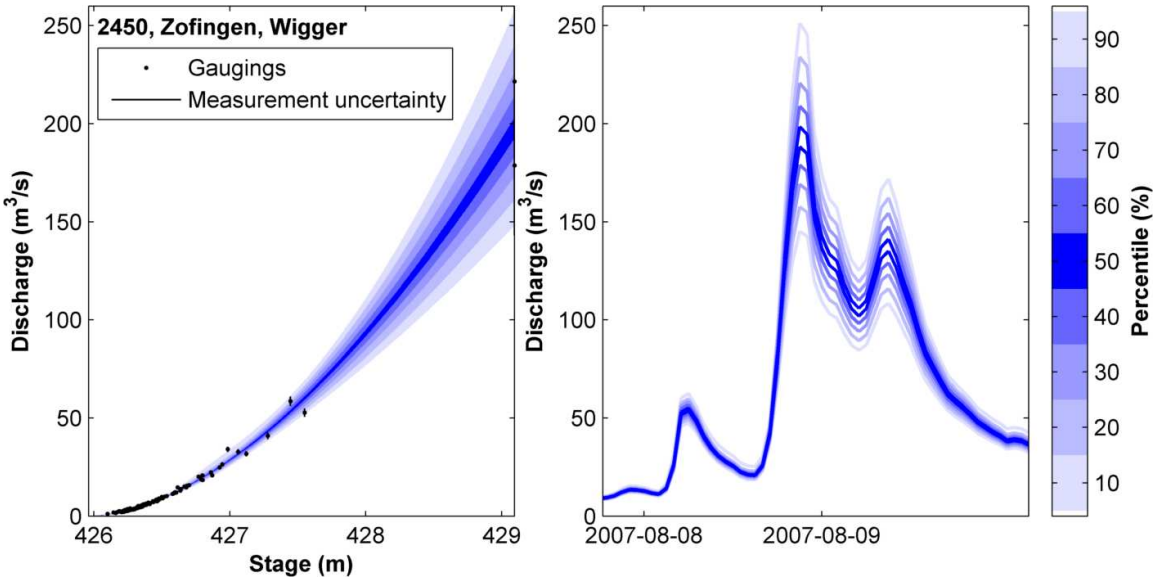


Figure 112 *Rating curve for Wigger at Zofigen (left) and flood event of August 2007 (right).*

6.15 Kander, Hondrich

Discharge uncertainty was estimated for 1980–2014. All the gauging data 1980–2016 were used for this station, there was no evidence about temporal change and the gaugings showed relatively little spread. There were several high flow gaugings and little extrapolation. A one section power law curve showed a good fit to the data (Figure 113).

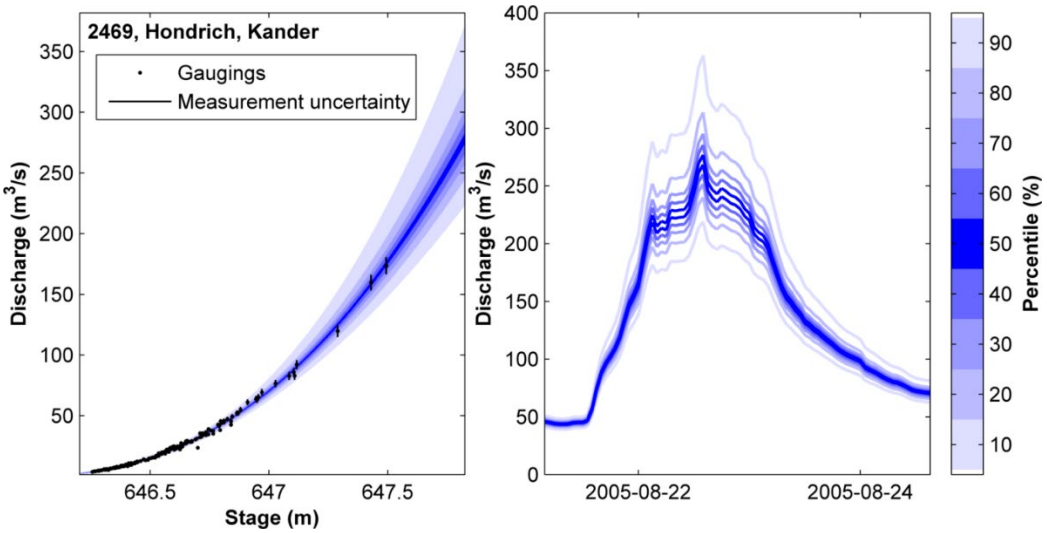


Figure 113 Rating curve for Kander at Hondrich (left) and flood event of August 2005 (right).

6.16 Murg, Murgenthal

Discharge uncertainty was estimated for 1980–2014. All the gauging data until May 2015 were used. There was a large change to the rating at the 2007 flood event where the river bed was lowered about 30 cm. The gauging data were divided into three estimation periods: 1980-06-26–2006-04-11, 2006-04-11–2007-08-29, and 2007-08-29 to 2015-05-03. A two-section power law curve was used in each period. The estimations were only made up to the bank-full level of 419.9 m. The rating curve for the latest period and the flood peak for the August 2007 event are shown below.

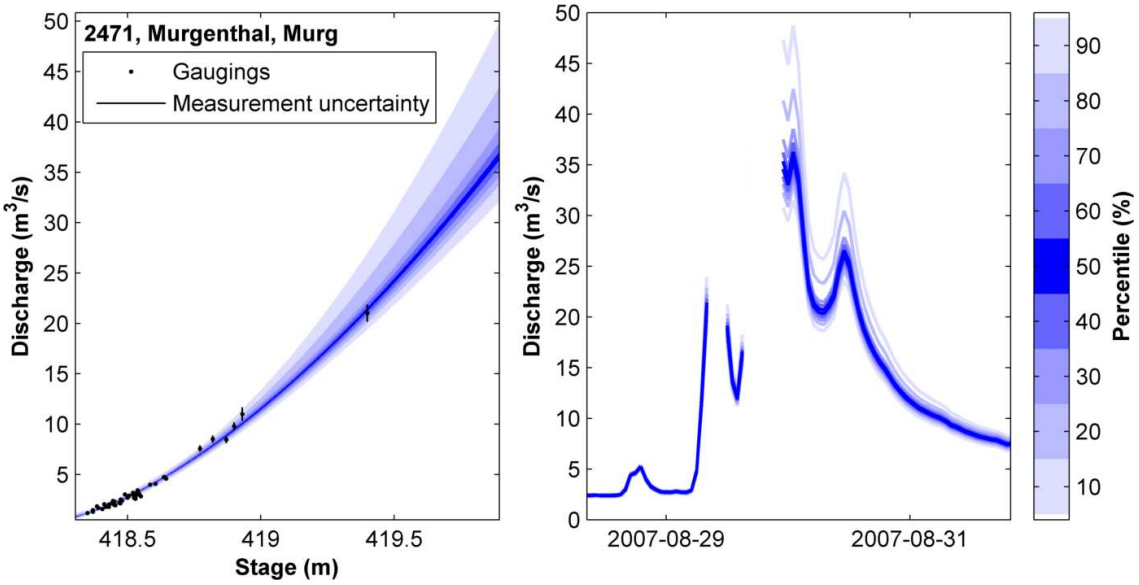


Figure 114 Rating curve for Murg, Murgenthal (left) and flood event of August 2007 (right).

The rating curve changed during the 2007 flood event so that the first part of the flood event is estimated with the rating curve for the first period (Figure 115, left), and the second part with the latest rating curve period (Figure 114, left). The rating curve for the earliest period is show in Figure 115 (right).

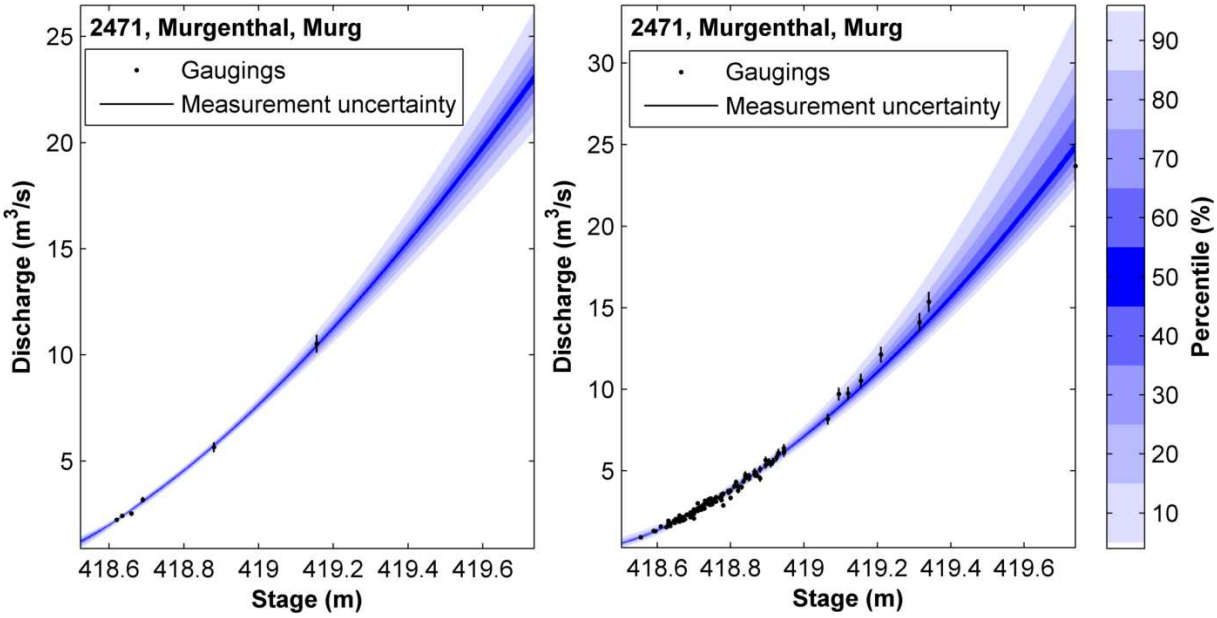


Figure 115 Rating curves resulting from data of the period 1980-06-26 to 2006-04-11 (left) and data of the period 2007-08-29 to 2015-05-03 (right).

6.17 Simme, Latterbach

There was a large change in the rating during the 2005 floods at this station. Therefore two different rating curves were used, for the periods 1990-01-01 to 2005-08-22 and 2007-04-07 to 2014-12-31. For the period in-between the 2005 and 2007 floods (2005-08-22 to 2007-04-07) there were large changes in the rating relation and not enough data for reliable estimation of uncertainty, this period was therefore excluded. A two-section power law curve was used for both periods. The rating curve for the earlier period and the 2005 flood event are shown below.

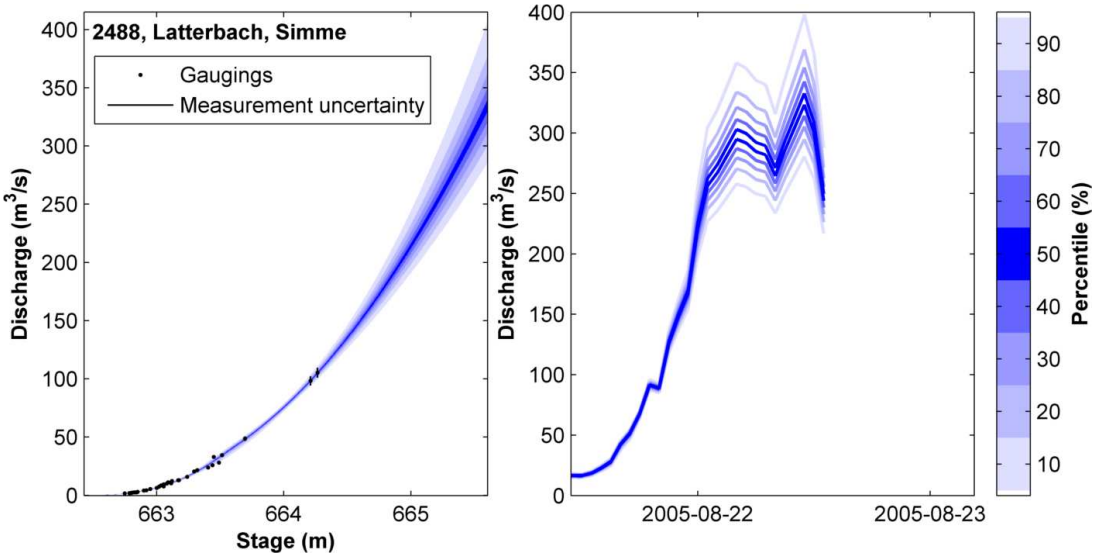


Figure 116 Rating curve for the Simmer at Latterbach and the flood event of August 2005.

The rating curve uncertainty estimated for the later period is shown in Figure 117.

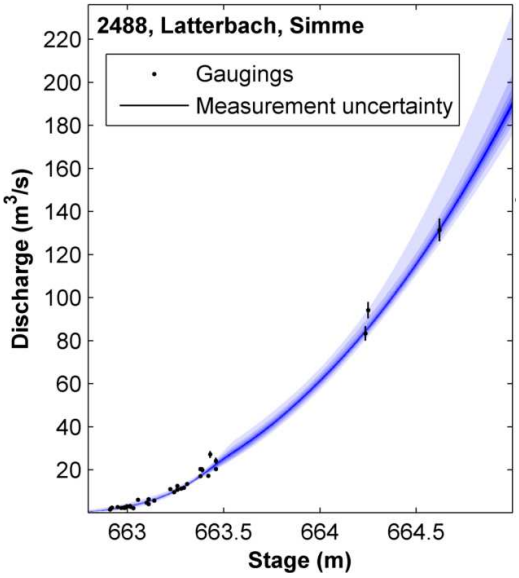


Figure 117 Rating curve for Simme, Latterbach using the period 2007-04-07 to 2014-12-31.

7 Jura lakes and the operation of the weir Port in RS Minerve

Martina Kauzlaric

In RS Minerve's representation of the total Aare River system, the most complex and most important lake is probably Lake Biel. This is due to its strong interaction with the two neighboring lakes upstream, namely Lake Neuchâtel and Lake Morat, in combination with its regulation at the weir in Port. Owing to this particular set-up, Lake Biel can have a strong attenuating effect on floods in the Aare River. For this reason, and for a better understanding of the results shown in the main report, this chapter explains in more detail how the interactions between the three Jura lakes and the Port Weir were implemented in RS Minerve.

Lake Morat and Lake Neuchâtel are connected by the Broye Canal, while Lake Neuchâtel and Lake Biel are connected further downstream by the Zihl Canal. In both of these canals, water can flow in both directions, depending on the stages of the lakes. The discharge in the canals is determined by the magnitude of stage differences. Relationships between stage, stage difference and discharge were initially implemented in RS Minerve by digitizing data from a report on extreme floods in the Aare River by geo7 et al. (2007). The data in this report were in turn extrapolated by means of 1D stationary hydraulic simulations. The digitized relationships were refined with help of linear interpolation, in order to have a rules with an interval of 6.25 cm stage. This resulted in 26 rules for the Broye Canal and 49 for the Zihl Canal, for both in each flow direction. The outflow of Lake Biel is regulated at the Port Weir in Nidau, and the rules for this regulation depend on the water level in the lake as well as on the day of the year. These rules were summarized at a monthly time-step, as shown in Figure 118. Additional rules have been implemented taking into account the so-called "Murgenthaler Bedingung" (Murgenthal constraint), according to which discharge at Murgenthal mustn't exceed 850 m^3/s, as well as considering the possibility of either curbing lake outflow in case of retention potential left in the lake (stage 429.6–429.8 m a.s.l., discharge at Murgenthal 500–850 m^3/s), or increasing outflow in case of a high lake stage (stage 429.6–430.35 m a.s.l., discharge at Murgenthal 400–700 m^3/s). Between

lake stages of 430.2 m a.s.l. and 431.0 m a.s.l., all monthly rules linearly converge towards the capacity of the fully open weir, while between 431 and 432 m a.s.l., the water releases coincide with the fully open capacity of the Port Weir (see Figure 118). The flood routing in RS Minerve was calibrated with results from the hydraulic 2D modelling with BASEMENT v3, as was the discharge flowing downstream of the Port Weir with lake stages above 432 m a.s.l. (see red dashed line Figure 118). The threshold at which the water level reaches the elevation of the surroundings of the weir (around 434 m a.s.l.) is reflected in the change from a linear relathionship to an exponential one.

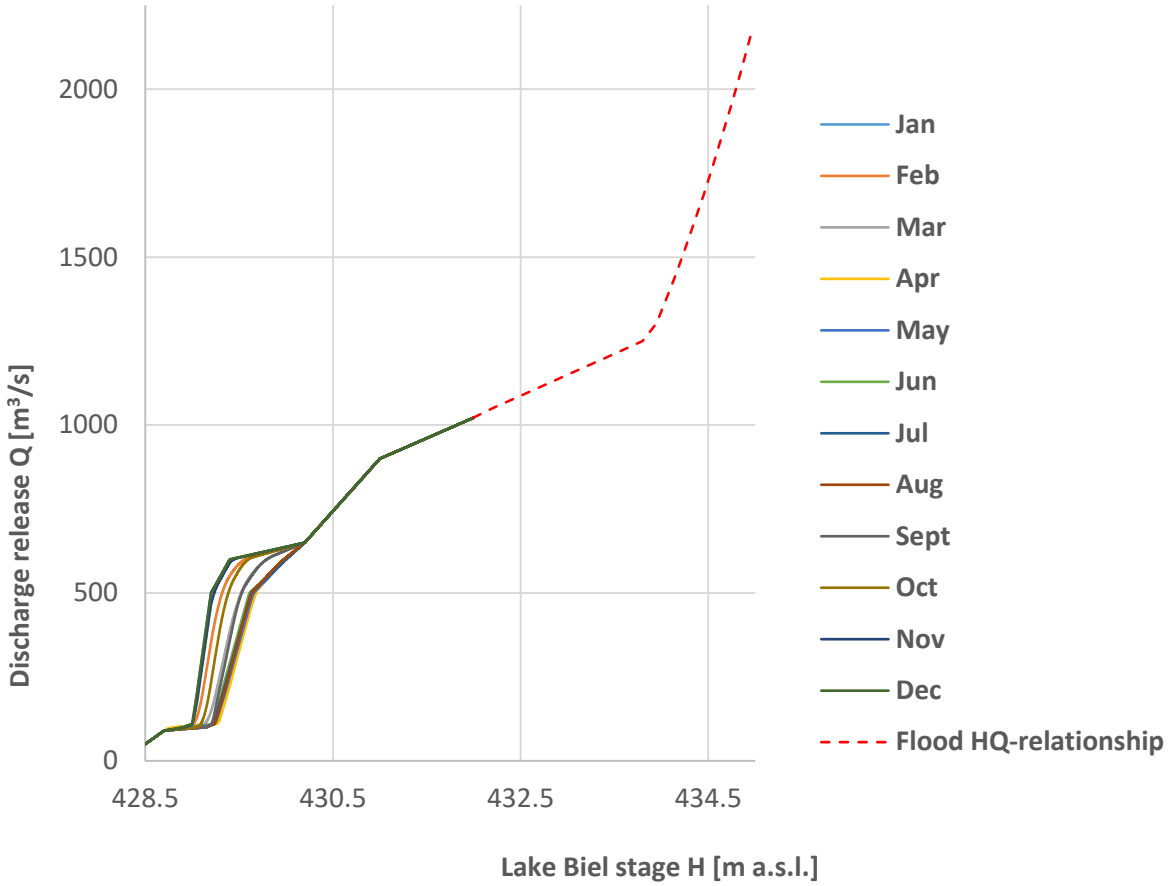


Figure 118 Operating rules applied at the outlet of Lake Biel in RS Minerve: twelve monthly rules for operation under normal to high-water conditions (*colored lines*), and under flood conditions (*dashed red line*).

The interaction of the three lakes and the weir has potentially a quite strong attenuating effect. This has been observed in 2005, when more than 450 m^3/s flowed back towards Lake Neuchâtel. The brilliant engineering of the system is also effective for extreme events, as can be seen in the plateau apparent for the outflow of Lake Biel in the simulation results of EXAR.

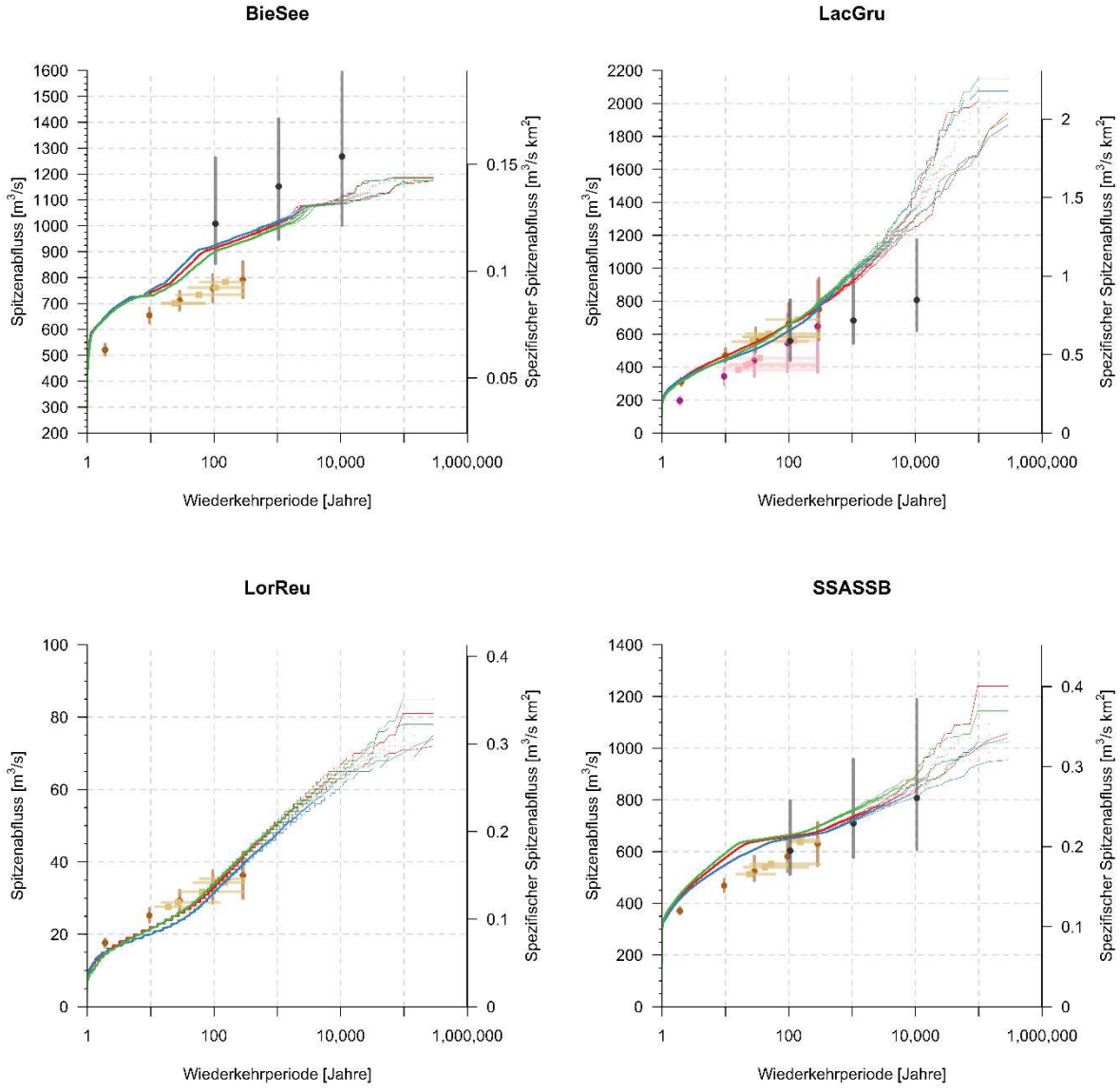
8 Exceedance curves for all transfer points

Daniel Viviroli, Calvin Whealton

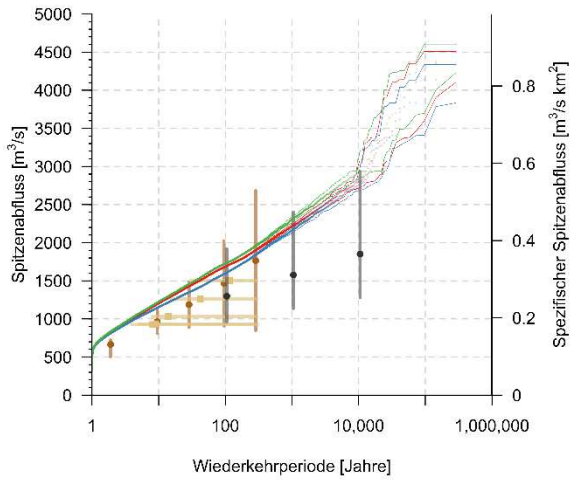
The following Figure 119 shows exceedance curves for all transfer points (Appendix 10.1, Table 2) as well as for additional important points represented in the model for the total Aare River system. In addition to the simulated annual maximum floods (semi-transparent points), the figures contain the following data (for further details see Table 5):

- „BAFU Intervalle“: Estimates of 2-, 10-, 30-, 100-, and 300-yr floods and associated intervals, based on gauged discharge and the methods described in Baumgartner et al. (2013). These values correspond to extrapolations done by the Federal Office for Environment (FOEN).
- „BAFU 5 grösste Spitzen“: Top 5 floods on record (FOEN measurement network) with intervals estimated from the uncertainty of the fitted distribution displayed in the plot. The green diamonds represent a bound of 300 yr return period, beyond which an extrapolation was not feasible.
- „EPFL Intervalle“: Estimates of the 100-, 1,000-, and 10,000-yr floods with prediction intervals according to Asadi et al. (2018).
- Estimates for historical floods where available.

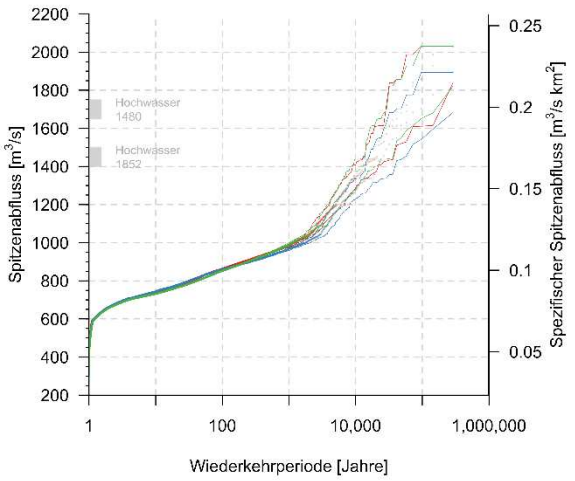
The uncertainty bands of simulated annual maximum floods were generated by bootstrap resampling with replacement and are intended to be interpreted as a 95% interval. Because bootstrap resampling cannot generate values beyond the original series, the narrowing of the interval towards the high floods is somewhat artificial.



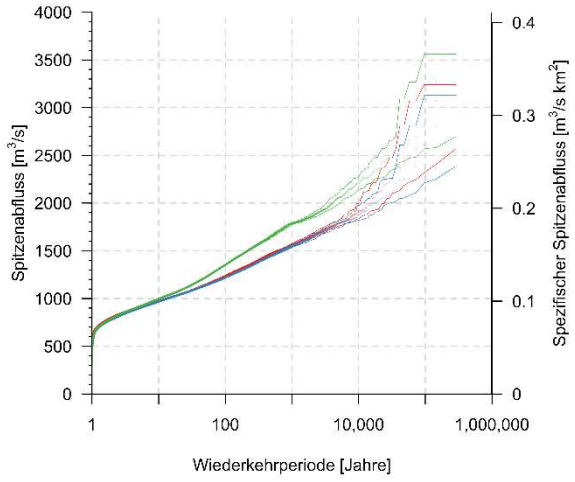
SSBSSC



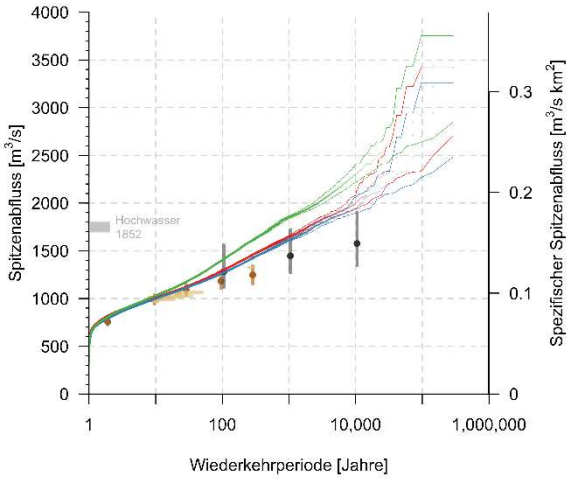
SSCSSD



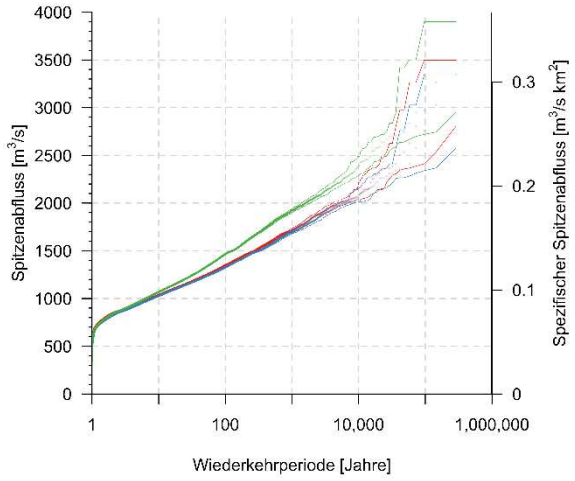
SSDSSE



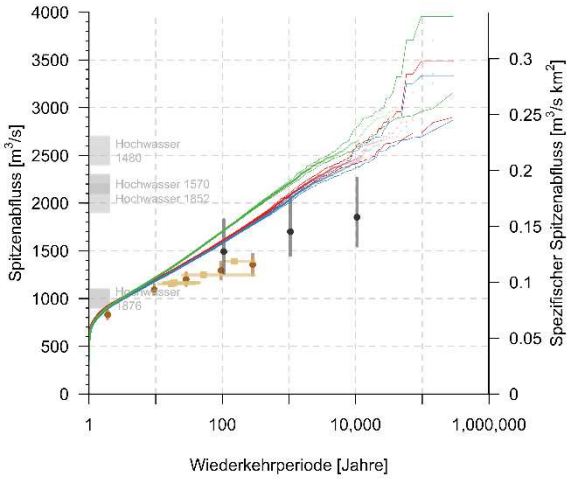
SSESSF



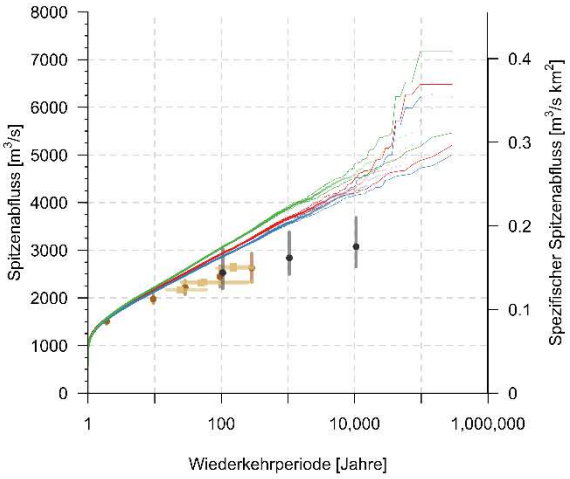
SSFSSG



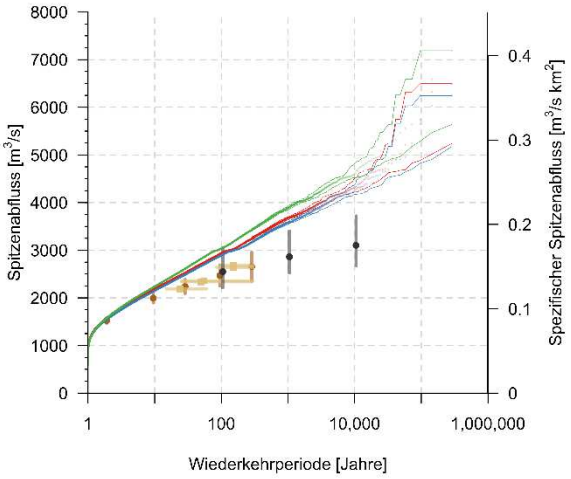
SSGSSH



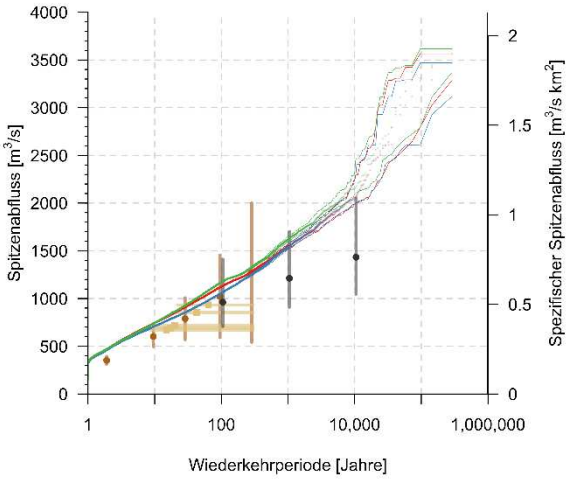
SSHSSI



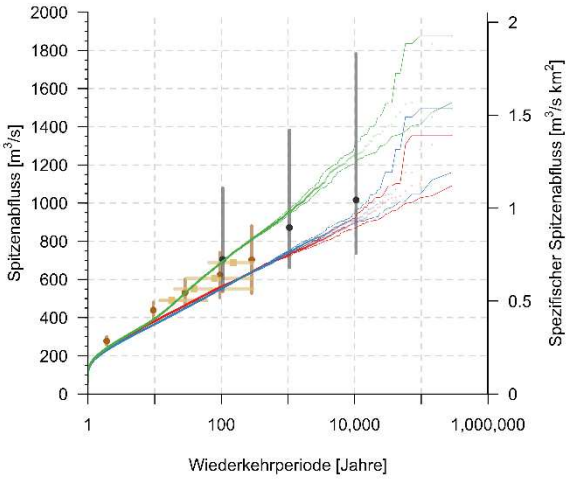
SSIOut



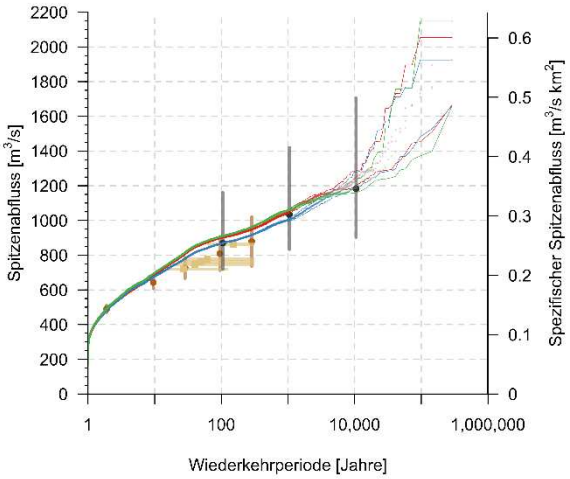
SSJSSB



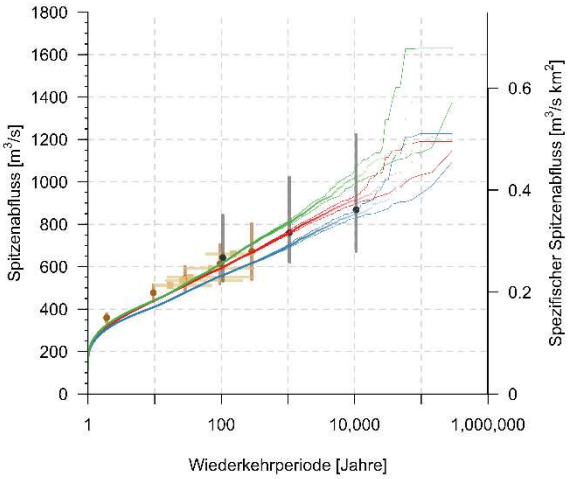
SSKSSD



SSLSSH



SSMSSH



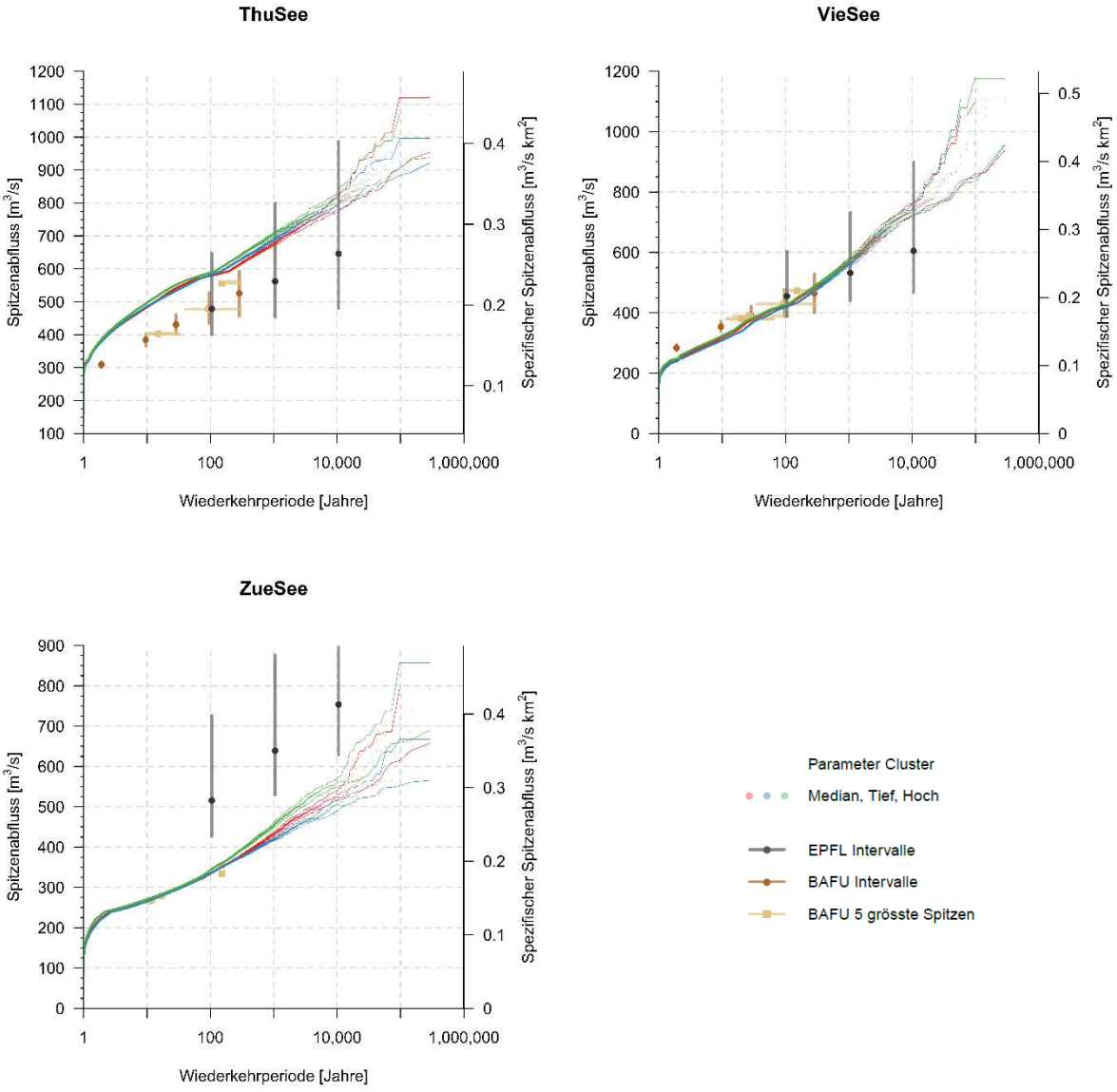


Figure 119 Exceedance plots for all transfer points and additional important points in the total system model.

Table 9 Summary of all return periods for all TPs.

	Return Period Floods Mean (2.5%-97.5%) [m^3/s]		
TP	**100-year**	**1000-year**	**10000-year**
LacGru	643 (615-660)	953 (905-993)	1354 (1258-1469)
LorReu	33 (31-34)	50 (47-51)	64 (61-66)
SSASSB	657 (649-664)	736 (714-762)	862 (816-888)
SSBSSC	1666 (1592-1729)	2226 (2137-2323)	2826 (2694-2983)
SSCSSD	859 (851-864)	981 (961-998)	1328 (1243-1409)
SSDSSE	1267 (1216-1358)	1628 (1540-1793)	1991 (1826-2263)
SSESSF	1321 (1263-1409)	1703 (1601-1864)	2086 (1905-2374)
SSFSSG	1381 (1323-1471)	1777 (1684-1932)	2181 (2015-2456)
SSGSSH	1630 (1575-1710)	2107 (2012-2222)	2522 (2378-2747)
SSHSSI	2946 (2849-3061)	3708 (3546-3911)	4392 (4164-4763)
SSIOut	2956 (2886-3043)	3714 (3553-3918)	4404 (4175-4754)
SSJSSB	1116 (1050-1173)	1566 (1517-1641)	2106 (1989-2291)
SSKSSD	604 (550-692)	812 (720-958)	1028 (874-1268)
SSLSSH	891 (866-908)	1035 (998-1062)	1219 (1163-1282)
SSMSSH	588 (556-617)	753 (688-812)	924 (831-1036)
ThuSee	583 (579-589)	690 (671-711)	802 (775-826)
VieSee	424 (418-429)	567 (554-579)	739 (722-762)
ZueSee	337 (333-344)	435 (416-460)	530 (493-566)
BieSee	911 (897-924)	1005 (987-1019)	1094 (1083-1106)

Table 10 Details for the data represented in the exceedance plots (Figure 119).

Site	**FOEN gauge**	**FOEN gauge area [km²]**	**Years of historical floods considered**	**Notes**
SSASSB	2135: Aare-Bern, Schönau	2‘941	-	
SSBSSC	2085: Aare-Hagneck	5‘112	-	
SSCSSD	-	-	1480; 1852	
SSDSSE	-	-	-	
SSESSF	2063: Aare-Murgenthal	10‘059	1852	
SSFSSG	-	-	-	
SSGSSH	2016: Aare-Brugg	11‘681	1480; 1507; 1852; 1876	
SSHSSI	2205: Aare-Untersiggenthal, Stilli	17‘553	-	
SSIOut	2205: Aare-Untersiggenthal, Stilli	17‘553	-	[1]
SSJSSB	2215: Saane-Laupen	1‘862	-	
SSKSSD	155: Emme-Wiler, Limpachmünd.	937	-	
SSLSSH	2018: Reuss-Mellingen	3‘386	-	
SSMSSH	2043: Limmat-Baden, Limmatpr.	2‘394	-	
BieSee	2029: Aare-Bruegg, Aergerten	8‘249	-	
LacGru	2119: Saane-Freiburg 2160: Sarine-Broc, Chateau d’en bas	1‘271 636	-	[2]
LorReu	2125: Lorze-Frauenthal	262	-	
ZueSee	--	-	-	[3]
VieSee	2152: Reuss-Luzern, Geissmattbr.	2‘254	-	
ThuSee	2030: Aare-Thun	2‘459	-	

[1] No major tributaries between SSHSSI and SSIOut, so the same gauge is used.
[2] EPFL estimates are for lake outflow. FOEN gauges are directly upstream or substantially downstream of the lake outlet, so they can only be compared on the specific discharge axis. Series 1 is Gauge 2119 and Series 2 is 2160 in the figure. The area for LacGru is 954 km².
[3] Only 5 peak floods calculated as difference of 2099 and 2176 gauges [area = 2174 -343 km²]

9 References

Asadi P., Engelke S., Davison A.C. (2018) Optimal regionalization of extreme value distributions for flood estimation. Journal of Hydrology, 556, 182–193. doi:10.1016/j.jhydrol.2017.10.051.

Bárdossy, A., & Pegram, G. G. S. (2009). Copula based multisite model for daily precipitation simulation. *Hydrology and Earth System Sciences*, *13*(12), 2299-2314.

Baumgartner E, Boldi M-O., Kan C., Schick S. (2013) Hochwasserstatistik am BAFU – Diskussion eines neuen Methodensets. Wasser, Energie, Luft, 105(2), 103–110.Bárdossy, A. (2007). Calibration of hydrological model parameters for ungauged catchments. Hydrol. Earth Syst. Sci., 11, 703–710, www.hydrol-earth-syst-sci.net/11/703/2007/

Chardon, J., Favre, A. C., & Hingray, B. (2016). Effects of spatial aggregation on the accuracy of statistically downscaled precipitation predictions. *Journal of Hydrometeorology*, *17*(5), 1561-1578.

Chardon, J., Hingray, B., & Favre, A. C. (2017). An adaptive two-stage analog/regression model for probabilistic prediction of local precipitation in France. *Hydrol. Earth Syst. Sci. Discuss., https://doi. org/10.5194/hess-2017-62, in review*.

CHy. 2010 (2010) Report on Conference on “Operational Hydrological Forecasting.” Bern: Swiss Hydrological Commission CHy of the Swiss Academy of Sciences.

Clark, M., Gangopadhyay, S., Hay, L., Rajagopalan, B., & Wilby, R. (2004). The Schaake shuffle: A method for reconstructing space–time variability in forecasted precipitation and temperature fields. *Journal of Hydrometeorology*, *5*(1), 243-262.

Coles, S., Heffernan, J., & Tawn, J. (1999). Dependence measures for extreme value analyses. *Extremes*, *2*(4), 339-365.

Davison, A. C., Huser, R., & Thibaud, E. (2013). Geostatistics of dependent and asymptotically independent extremes. *Mathematical Geosciences*, *45*(5), 511-529.

Evin, G., Blanchet, J., Paquet, E., Garavaglia, F., & Penot, D. (2016). A regional model for extreme rainfall based on weather patterns subsampling. *Journal of Hydrology*, *541*, 1185-1198.

Evin, G., Favre, A. C., & Hingray, B. (2018). Stochastic generation of multi-site daily precipitation focusing on extreme events. *Hydrology and Earth System Sciences*, *22*(1), 655-672.

Evin, G., Favre, A. C., & Hingray, B. (2019). Stochastic generators of multi-site daily temperature: comparison of performances in various applications. *Theoretical and Applied Climatology*, 1-14.

Fallot, J.M. (2017). PMP (Probable Maximum Precipitation) Maps for Switzerland". Verification of flood safety for dams: Application of the CRUEX++ methodology for the estimation of extreme floods. CRUEX workshop, March 2017, Lausanne.

Favre, A. C., Quessy, J. F., & Toupin, M. H. (2018). The new family of Fisher copulas to model upper tail dependence and radial asymmetry: Properties and application to high-dimensional rainfall data. *Environmetrics*, *29*(3), e2494.

Felder G. & Rolf Weingartner R. (2016) An approach for the determination of precipitation input for worst-case flood modelling. *Hydrological Sciences Journal*, 61:14, 2600-2609. doi: 10.1080/02626667.2016.1151980

Felder G., Weingartner R. (2017) Assessment of deterministic PMF modelling approaches. *Hydrological Sciences Journal*, 62:10, 1591-1602. doi: 10.1080/02626667.2017.1319065

Felder G., Paquet E., Penot D., Zischg A., Weingartner R. (2019) Consistency of Extreme Flood Estimation Approaches. *J. Hydrol. Eng.*, 24(7): 04019018.doi: 10.1061/(ASCE)HE.1943-5584.0001797

geo7 (2007). Extremhochwasser im Einzugsgebiet der Aare, EHW Aare, Bericht; geo7, IUB, HZP, Emch+Berger AG

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1-2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Hertig, J. A., & Fallot, J. M. (2009). Project Cruex. *Directive crue de l'OFEN. Carte de précipitations extrêmes PMP pour la Suisse. Validation et utilisation des cartes de PMP pour l'obtention de la PMF. Hertig & Lador*.

Keller, D. E., Fischer, A. M., Frei, C., Liniger, M. A., Appenzeller, C., & Knutti, R. (2015). Implementation and validation of a Wilks-type multi-site daily precipitation generator over a typical Alpine river catchment. *Hydrology and Earth System Sciences*, *19*(5), 2163-2177.

Kienzler P., Andres N., Näf-Huber D., Zappa M. (2015) Herleitung extremer Niederschläge und Hochwasser im Einzugsgebiet des Sihlsees für einen verbesserten Hochwasserschutz der Stadt Zürich. *Hydrologie und Wasserbewirtschaftung HyWa*, 59, 48-58. doi: 10.5675/HyWa_2015,2_1

McMillan, H. K. & Westerberg, I. K. (2015). Rating curve estimation under epistemic uncertainty. Hydrological Processes 29: 1873–1882, doi: 10.1002/hyp.10419.

McNeil, A. J., Frey, R., & Embrechts, P. (2005). *Quantitative risk management: Concepts, techniques and tools* (Vol. 3). Princeton: Princeton university press.

Mezghani, A., & Hingray, B. (2009). A combined downscaling-disaggregation weather generator for stochastic generation of multisite hourly weather variables over complex terrain: Development and multi-scale validation for the Upper Rhone River basin. *Journal of Hydrology*, *377*(3-4), 245-260.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Naveau, P., Huser, R., Ribereau, P., & Hannart, A. (2016). Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. *Water Resources Research*, *52*(4), 2753-2769.

Papastathopoulos, I., & Tawn, J. A. (2013). Extended generalised Pareto models for tail estimation. *Journal of Statistical Planning and Inference*, *143*(1), 131-143.

Poli, P., Hersbach, H., Dee, D.P., Berrisford, P., Simmons, A.J., Vitart, F., Laloyaux, P., Tan, D.G., Peubey, C., Thépaut, J., Trémolet, Y., Hólm, E.V., Bonavita, M., Isaksen, L., & Fisher, M. (2016). ERA-20C: An atmospheric reanalysis of the twentieth century. *Journal of Climate*, *29*(11), 4083-4097.

Pool S., Vis M. & Seibert J. (2018) Evaluating model performance: towards a non-parametric variant of the Kling-Gupta efficiency, *Hydrological Sciences Journal*, 63:13-14, 1941-1953, DOI: 10.1080/02626667.2018.1552002

Raynaud, D., Hingray, B., Zin, I., Anquetin, S., Debionne, S., & Vautard, R. (2017). Atmospheric analogues for physically consistent scenarios of surface weather in Europe and Maghreb. *International Journal of Climatology*, *37*(4), 2160-2176.

Raynaud, D., & Hingray, B. (2019). Generating new atmospheric trajectories. *In preparation*.

Schälchli, Abegg + Hunzinger AG (2007). Gefahrenkarte Hochwasser Wiggertal, Gemeinden Brittnau, Strengelbach und Zofingen. Technischer Bericht und Massnahmenplanung. Zürich.

Wilks, D. S. (1998). Multisite generalization of a daily stochastic precipitation generation model. *Journal of Hydrology*, *210*(1-4), 178-191.

Westerberg, I. K., Wagener, T., Coxon, G., McMillan, H. K., Castellarin, A., Montanari, A., Freer, J. (2016). Uncertainty in hydrological signatures in gauged and ungauged catchments. Water Resources Research, 52, 1847–1865, doi:10.1002/2015WR017635.

WMO (World Meteorological Organization) 2009. Manual on estimation of probable maximum precipitation (PMP). Geneva: WMO.

Viviroli D, Zappa M, Gurtz J, Weingartner R (2009) An introduction to the hydrological modelling system PREVAH and its pre- and post-processing tools. Environmental Modelling & Software, 24(10), 1209–1222, doi:10.1016/j.envsoft.2009.04.001

Zischg A. P., Felder G., Weingartner R., Quinn N., Coxon G., Jeffrey N., Freer J., Bates P. (2018) Effects of variability in probable maximum precipitation patterns on flood losses. *Hydrol. Earth Syst. Sci.,* 22, 2759–2773. https://doi.org/10.5194/hess-22-2759-2018

10 Appendixes

10.1 IDs of catchments and transfer points

Table 11 Explanation of IDs for HBV catchments

ID	River and site
AabLen	Aabach-Lenzburg
AabSch	Aabach-Schmerikon
AarBri	Aare-Brienzwiler
ArbAve	Arbogne-Avenches
AreBou	Areuse-Boudry
ArnGra	Arnon-Grandson
BibKer	Bibere-Kerzers
BroPay	Broye-Payerne
BueOth	Bünz-Othmarsingen
ChaAve	Chandon-Avenches
ChiFre	Chise-Freimettingen
ChiXXX	Chise-Restgebiet
DueOlt	Dünnern-Olten
EBABuo	Engelberger Aa-Buochs
EmmBur	Emme-Burgdorf
FurWue	Furtbach-Würenlos
GalTaf	Galtera / Gottéron-Tafers
GieBri	Giessbach-Brienz
GlaXXX	Glâne
GroIse	Grosstalbach-Isenthal
GueBel	Gürbe-Belp
JogBro	Jogne-Broc
JonRue	Jona-Rüti
KanHon	Kander-Hondrich
KEmEmm	Kleine Emme-Emmen
LinWee	Linth-Weesen
LorZug	Lorze-Zug
LueGst	Lütschine-Gsteig
MenYvo	Mentue-Yvonand
MuoIng	Muota-Ingenbohl
MurMur	Murg-Murgethal
NozPrC	Nozon-Pré Chaillet
OenHei	Önz-Heimenhusen
OesKri	Ösch-Kriegstetten
OrbOrb	Orbe-Orbe
PGlVlG	Petite Glâne-Villars-le-Grand
RepDie	Reppisch-Dietikon
ReuSee	Reuss-Seedorf
RtaXXX	Rotache
SaASar	Sarner Aa-Sarnen
SarBro	Sarine-Broc
SenLau	Sense-Laupen
SeyVal	Seyon-Valangin
SihZue	Sihl-Zürich

SimLat	Simme-Latterbach
SioVui	Sionge-Vuippens
SteKal	Steinenbach-Kaltbrunn
SuhUnt	Suhre-Unterentfelden
SurDoe	Surb-Döttingen
SuzPer	Suze-Péry
TalCha	Talent-Chavornay
WaegXX	Wägitaler Aa
WigZof	Wigger-Zofingen
WorItt	Worble-Ittigen
WynSuh	Wyna-Suhr
XBIXXX	Bielersee-Zuflüsse
XBRXNW	Brienzersee-Zuflüsse Nordwest
XBRXSE	Brienzersee-Zuflüsse Südost
XGRXXX	Greyerzersee-Zuflüsse
XMUXXX	Murtensee-Zuflüsse
XNEXNW	Neuenburgersee-Zuflüsse Nordwest
XNEXSE	Neuenburgersee-Zuflüsse Südost
XTHXNE	Thunersee-Zuflüsse Nordost
XTHXSW	Thunersee-Zuflüsse Südwest
XVWXNE	Vierwaldstättersee-Zuflüsse Nordost
XVWXSW	Vierwaldstättersee-Zuflüsse Südwest
XZGXXX	Zugersee, Zuflüsse
XZHXNE	Zürichsee-Zuflüsse Nordost
XZHXNW	Zürichsee-Zuflüsse Nordwest
XZHXSE	Zürichsee-Zuflüsse Südost und Linthebene
XZHXSW	Zürichsee-Zuflüsse Südwest
YAAYY1	Aare Restgebiet 1 (Thun bis Bielersee)
YAAYY2	Aare Restgebiet 2 (Bielersee bis Emme)
YAAYY3	Aare Restgebiet 3 (Emme bis Rhein)
YEMYYY	Emme Restgebiet
YLIYYY	Limmat Restgebiet
YREYYY	Reuss Restgebiet
YSAYYY	Sarine Restgebiet
ZulXXX	Zulg

Table 12 Explanation of IDs for RS nodes.

ID	Transfer point and station
AarAeg	Aare at Aegerten
AarBru	Aare at Brugg
AarThu	Aare at Thun
AarUnt	Aare at Untersiggenthal
BieSee	Lake Biel (Total System)
EmmWil	Emme at Wiler
LacGru	Lake Gruyère (Total System)
LimBad	Limmat at Baden
LorReu	Lorze at confluence to Reuss
ReuLuz	Reuss at Luzern
ReuMel	Reuss at Mellingen
SaaLau	Saane at Laupen
ThuSee	Lake Thun (Total System)
VieSee	Lake Lucerne (Total System)
ZueSee	Lake Zurich (Total System)

10.2 Disaggregation of daily observations to an hourly resolution

EXAR

Task T3-GINP, Grenoble

Working report

July 29, 2016

Disaggregation of daily observations to an hourly resolution

Guillaume Evin, Jérémy Chardon,
Benoit Hingray and Anne-Catherine Favre

1 Introduction

Since the fine time scale resolution is absolutely necessary for the application of the hydraulic models, the meteorological scenarios – and consequently the hydrological scenarios – must be delivered at an hourly resolution. Daily weather scenarios (precipitation and temperature) generated by SCAMP and G-Wex must then be disaggregated to an hourly resolution. This report describes a first version of a disaggregation model which is based on an analog approach: for each target day and each site, the daily generated weather variables are disaggregated according to the temporal structure of this variable observed for an analogous day. Analog dates are identified using a distance criterion (e.g. the Root Mean Square Error, RMSE) which aims to measure the similarity between the observed and generated daily precipitation fields.
The disaggregation method can also be applied to daily observations and we present a short evaluation of the precipitation and temperature measurements disaggregated at an hourly scale.

2 Data

Predictors used to identify the analog dates correspond here to the daily precipitation field known over a set of S_d sites. Three datasets are used in the following, namely for any given day d:

- the simulated (possibly observed) daily precipitation fields $\tilde{\mathbf{P}}_d^{24} = \{\tilde{P}_{s,d}^{24}\}, s \in S_d$ from SCAMP or G-Wex over a set S_d of daily gauges. These simulated (or observed) weather variables are available on long periods (e.g. more than 100 years) and have to be disaggregated to an hourly resolution,
- The observed daily precipitation fields $\mathbf{P}_d^{24} = \{P_{s,d}^{24}\}, s \in S_d$ at the same set S_d of daily gauges. We restrict the archive to the time period where observed weather variables at an hourly resolution have been recorded, d is in the period 1990/01/01 - 31/12/2014,
- The observed hourly precipitation fields $\mathbf{P}_h^1 = \{P_{s,h}^1\}, s \in S_h$ at a set S_h of hourly gauges.

After application of the disaggregation approach, simulated (possibly observed) hourly precipitation fields $\tilde{\mathbf{P}}_h^1 = \{\tilde{P}_{s,h}^1, \ldots, \tilde{P}_{s,h}^1\}, s \in S_d$ at a set S_d of daily gauges will be obtained. Similarly to the precipitation fields $\mathbf{P}$, $\mathbf{T}$ will denote the temperature fields.

3 Description of the disaggregation method

3.1 Meteorological analog

Let note d, the target day to be disaggregated. The disaggregation of the daily precipitation amounts available for this day follows two steps:

1. A similarity criterion between the target day d and each candidate day d' is computed. Analog dates are selected using a distance between $\tilde{\mathbf{P}}_d^{24}$ and $\mathbf{P}_{d'}^{24}$ (observed and simulated daily precipitation fields for days d and d'). As a first approach, this distance is the RMSE (a small RMSE value indicating a strong similarity).

2. The best analog date is then selected.

3.2 Closest hourly stations as a proxy

For a simulated day d, let assume that an analog date d' has been found in the period 1990-01-01 to 2014-12-31. The simulated and observed precipitation fields at the daily gauges S_d are thus considered to be similar. For a given gauge s in S_d, we use the hourly temporal structure observed for the analog day d' at this station in order to disaggregate the daily amount of the target day at this station. When no hourly data is available at a given station in the analog day d', we use the temporal structure observed at the closest gauge s' in S_h (according to a crossing distance which penalizes the crossing of crests and valleys) where hourly data are available.

3.3 Temporal structure at an hourly scale

In case of precipitation, the temporal sub-daily structure is taken as the sequence of 24 ratios between the hourly amounts and the corresponding daily amount. The temporal structure for a day d is thus represented as a vector of 24 proportion values $\pi_{s,\mathbf{h}'}$ of the daily precipitation:

$$\pi_{s,\mathbf{h}} = P_{s',\mathbf{h}'}^{1} / P_{s,d'}^{24}, \tag{1}$$

where $\mathbf{h}$ and $\mathbf{h}'$ are the 24 hours that compose days d and d', respectively. The proportions $\pi_{s,\mathbf{h}}$ vary between 0 and 1 and $\sum_{\mathbf{h}} \pi_{s,\mathbf{h}} = 1$.
For temperature, the 24-value vector of the sub-daily structure is based on the following deviations:

$$\pi_{s,\mathbf{h}} = T_{s',\mathbf{h}'}^{1} - T_{s,d'}^{24} \tag{2}$$

between the hourly and the corresponding daily average temperatures.

3.4 Disaggregation of the generated scenarios

Finally, the hourly simulated scenarios are simply obtained as:

$$\tilde{P}_{s,\mathbf{h}}^{1} = \pi_{s,\mathbf{h}} \times \tilde{P}_{s,d}^{24}, \tag{3}$$

in the case of precipitation, and as:

$$\tilde{T}^{1}_{s,\mathbf{h}} = \pi_{s,\mathbf{h}} + \tilde{T}^{24}_{s,d}, \tag{4}$$

in the case of temperature. Figure 1 provides a summary of the whole procedure.

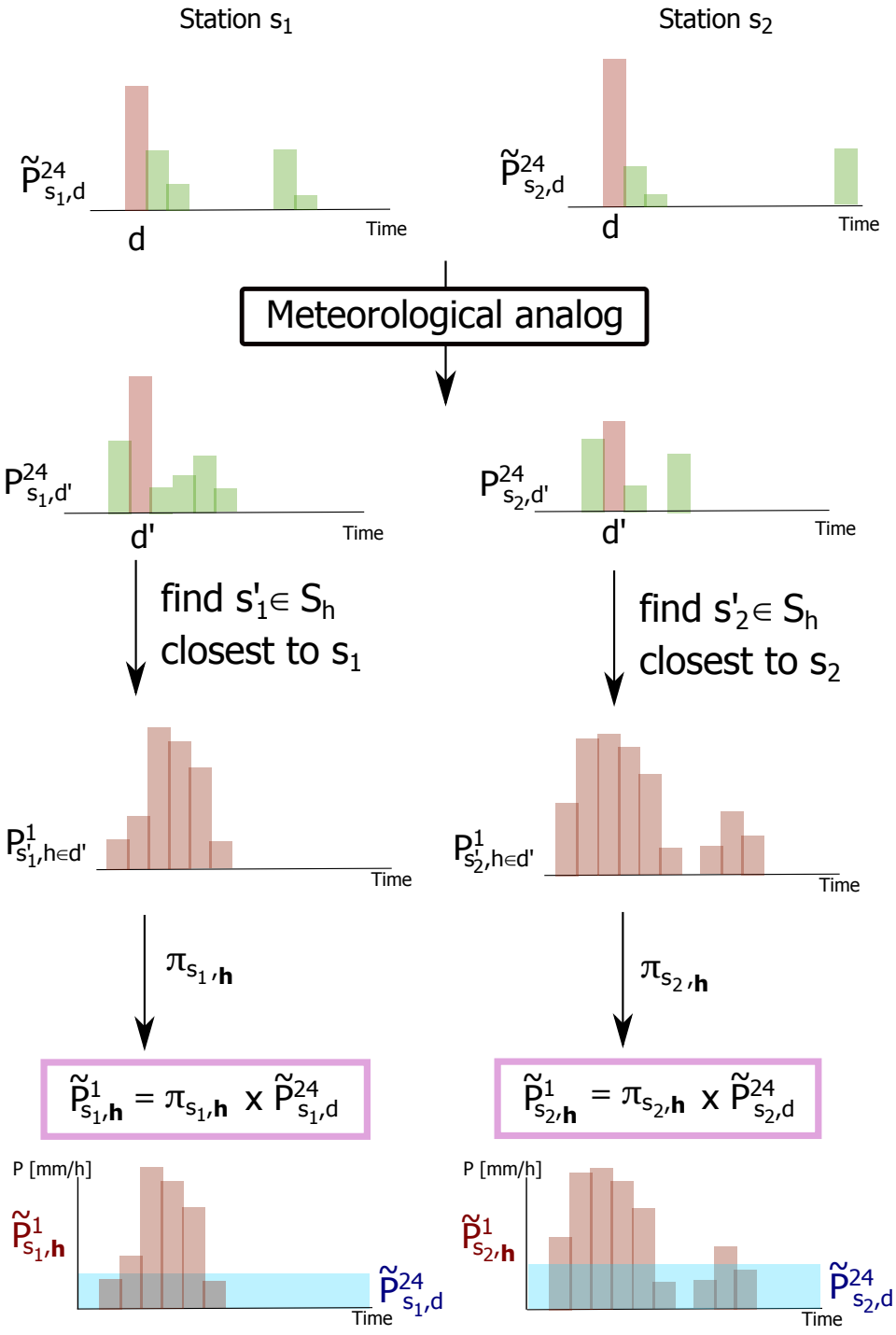


Figure 1: Illustration of the disaggregation method.

4 Application

In this section, we evaluate the properties of the daily observations disaggregated to an hourly scale using the method described above (i.e. $\tilde{\mathbf{P}}_d^{24}$ are observed daily fields of precipitation and temperature). The disaggregated values are compared to hourly observations.

4.1 Locations of the closest hourly stations

Because hourly data are usually not available at a given station where daily data has been measured, we use the temporal structure observed at the closest gauge s' in S_h (according to a crossing distance which penalizes the crossing of crests and valleys), where hourly data are available.
Figure 2 represents the 105 daily gauges in S_d and the 65 hourly gauges S_h selected for the disaggregation of precipitation data. For each daily gauge, a segment shows the closest hourly gauge according to the crossing distance. Since there are less hourly gauges than daily gauges in the study area, the same hourly gauge is often linked to several daily gauges, which means that the same sub-daily temporal structures will be attributed to these daily gauges.
Figure 3 represents the 26 daily gauges in S_d and the 67 hourly gauges S_h selected for the disaggregation of temperature data. Compared to precipitation data, less time series area are available during long periods (e.g. 1930-2014) and the density of hourly gauges is higher than daily gauges. Furthermore, for most of the daily gauges, hourly gauges are available at the same location.

4.2 Comparison of the daily observations disaggregated at an hourly scale and hourly observations

For the period 1930-1989, the observed daily fields have been disaggregated according to the method described above. An analog day is searched in the period 1990-2014 according to the meteorological analogy, and we retrieve the temporal structure from the closest hourly gauge.
For some stations, hourly precipitation data are available before 1990 at the same stations than the daily gauges. The differences between hourly observations and disaggregated values can thus be used to evaluate the effect of the meteorological analogy only.
Figure 4 represents the distributions of the non-zero differences of hourly precipitation values during the period 1970-1989. The boxes indicate that 50% of these differences lie between -0.5mm and 0.5mm, approximately, and the whiskers show that 95% of the differences are between -3mm and 3mm. Figure 5 shows the distributions of the differences for the temperature during the same period. We see that 50% of the differences are in the interval $[-2, 2]$ degrees, approximately, and that 95% of the differences are in the interval $[-4, 4]$ degrees.
Obviously, these differences can be larger, and can reach 40 mm for the precipitation and 15 degrees for the temperature. For the period 1990-2014, the meteorological analogy is not necessary, as we can directly obtain the temporal structure from the closest hourly gauge. For

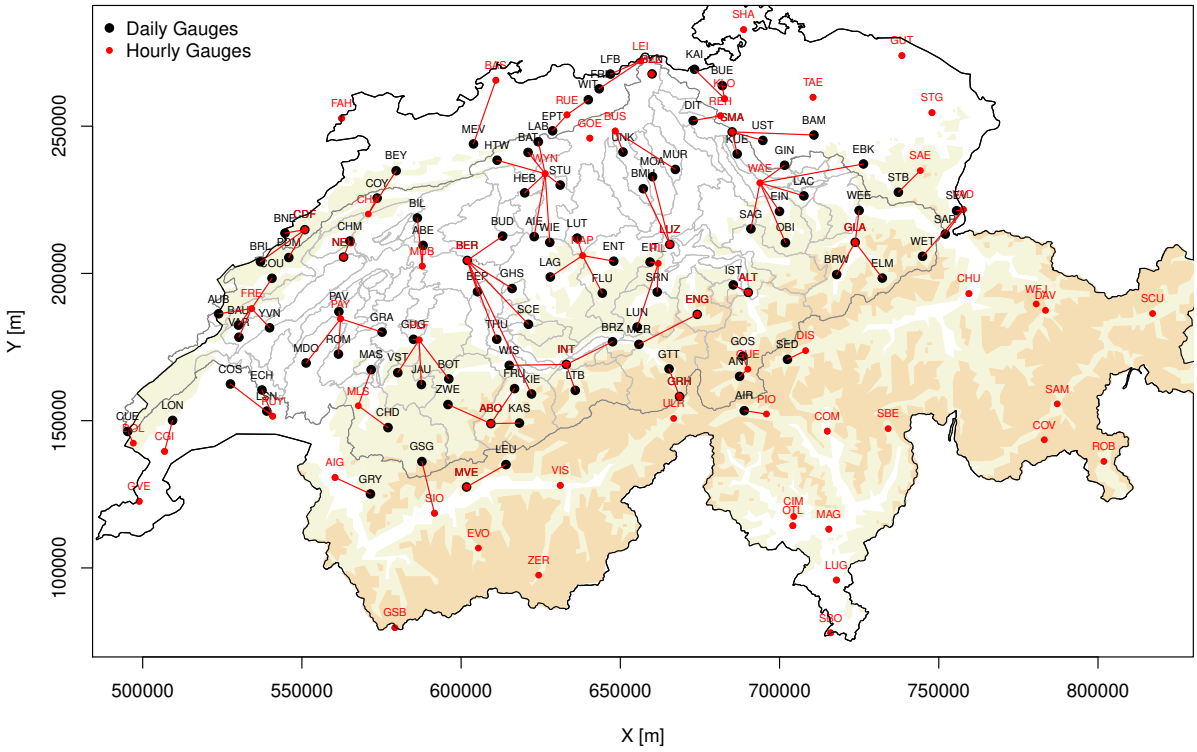


Figure 2: Hourly gauges taken as a proxy for the disaggregation of the precipitation simulated at the daily gauges.

the both precipitation and temperature, we select one station for which the differences are the largest, mainly due to the distance between the station where we apply the disaggregation method and the closest hourly gauge. The difference between the temporal structures thus indicate the impact of this distance.

Figure 6 shows the sub-daily temporal structures of observed hourly precipitation at the stations 'THU' at Thun ('Obs. 1H'), where daily observations are disaggregated, and 'BER' at Bern where hourly observations provide the temporal structure for the disaggregation ('Disag. 1H'). The two stations are located at a distance of approximately 28km. We can see that the largest differences often occur when there is an intense and short precipitation event. There is usually a lag of 1-2 hours between the peaks of these events at the two stations (e.g. on the 2013-07-18). Less frequently, the maximum intensities of this event are very different at the two locations (e.g. on the 2014-07-26).

Figure 7 shows the sub-daily temporal structures of temperature for stations 'LAG' at Langnau ('Obs. 1H') and 'NAP' at Napf ('Disag. 1H'), which are located at a distance of approximately 12km and at elevations of 745m and 1403m, respectively. For some days (e.g. '2013-12-02', '2012-02-08'), the temperatures have increased at Napf during the evening, which contrasts with the colder temperatures at Langnau. These higher temperatures at a higher elevation are explained by anticyclones over Switzerland which result in a stable weather. Warm air coming from North-East produces a sharp increase of the temperature at Napf, but not in the valley at Langnau. For other days (e.g. 2013-03-05, 2012-03-02),

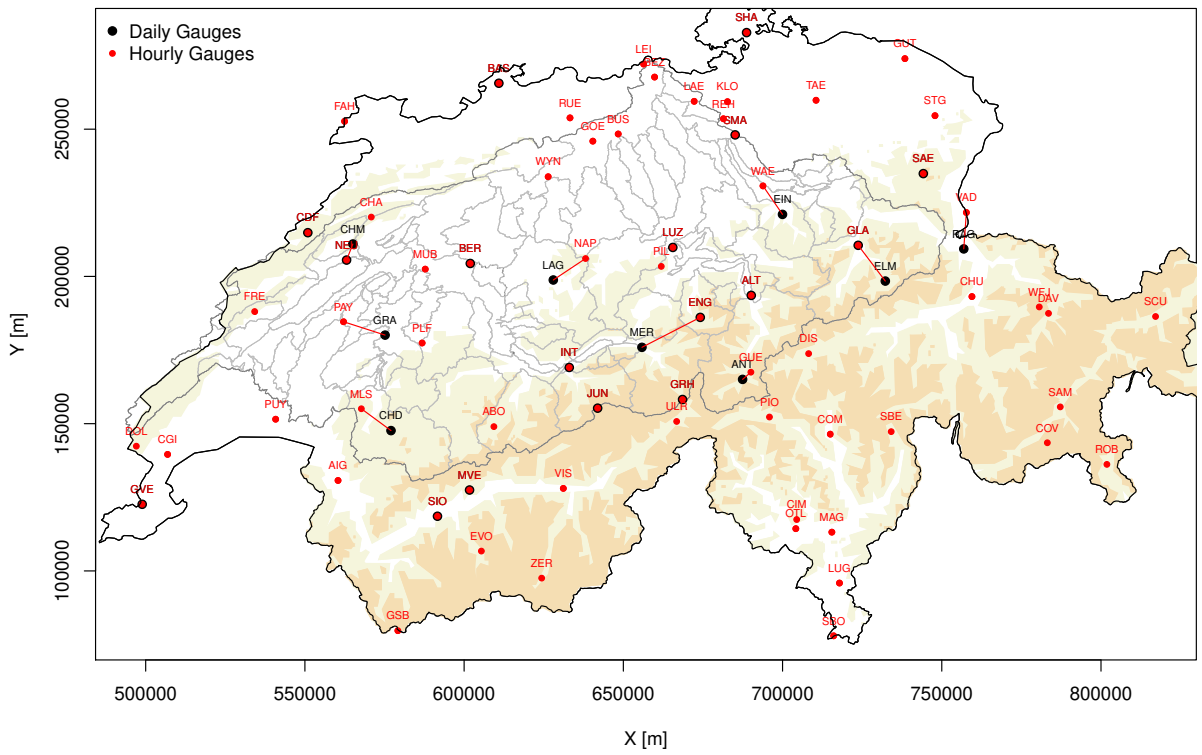


Figure 3: Hourly gauges taken as a proxy for the disaggregation of the temperature simulated at the daily gauges.

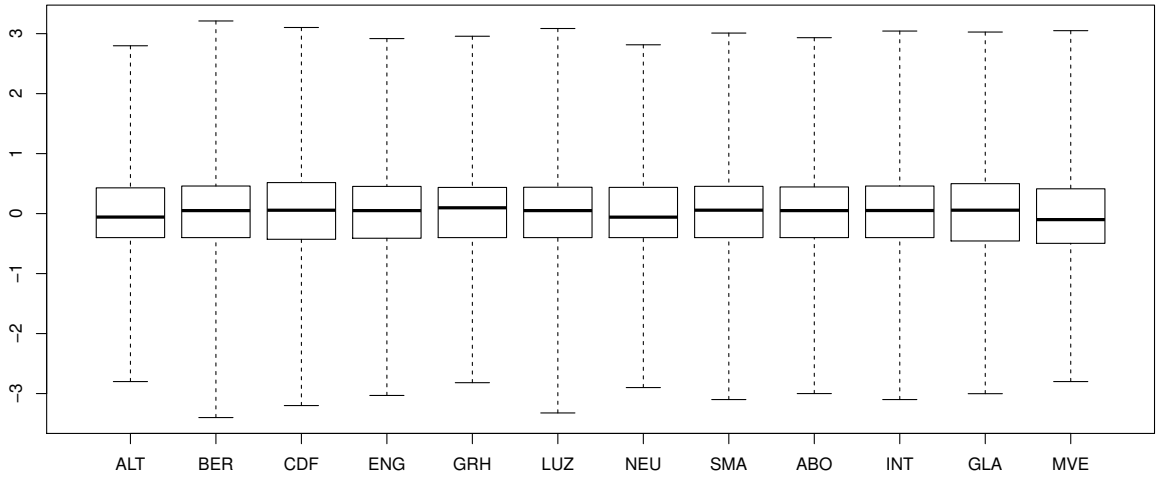


Figure 4: Boxplots of non-zero differences between disaggregated precipitation values and hourly observations for the period 1970-1989.

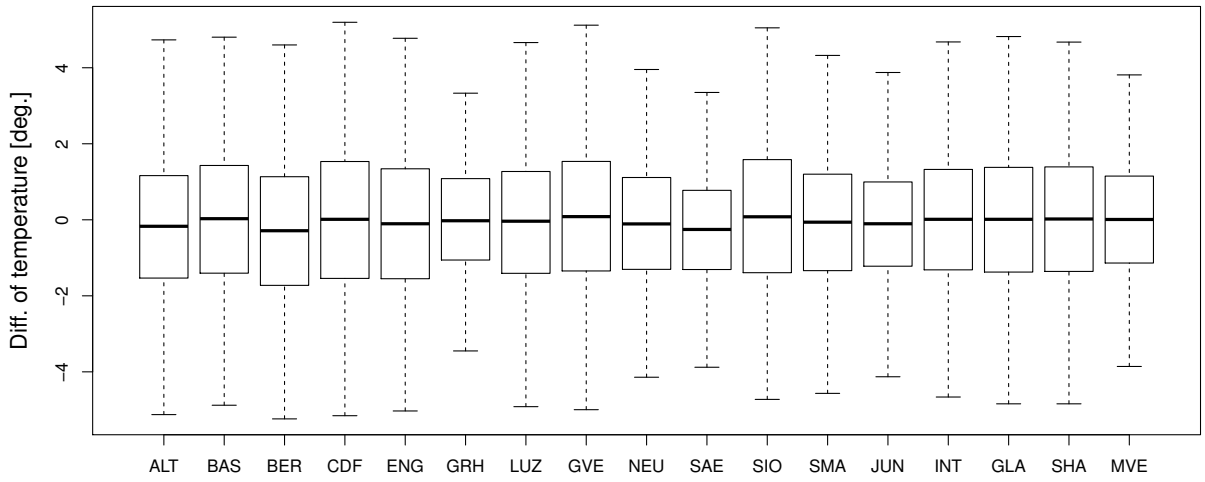


Figure 5: Boxplots of differences between disaggregated temperature values and hourly observations for the period 1970-1989.

the daily cycle of temperatures is very pronounced at Langnau, which is typical of stable atmospheric situations at the beginning of the spring. At higher elevations, this large differences of temperature between nights and days are not present, which explain why the temperatures are almost constant at Napf.
These differences in the sub-daily temporal structures at close locations clearly highlight the large spatial variations of precipitation and temperature for these fine time scale. The performances of the disaggregation method are thus inherently limited by the density of gauge networks.

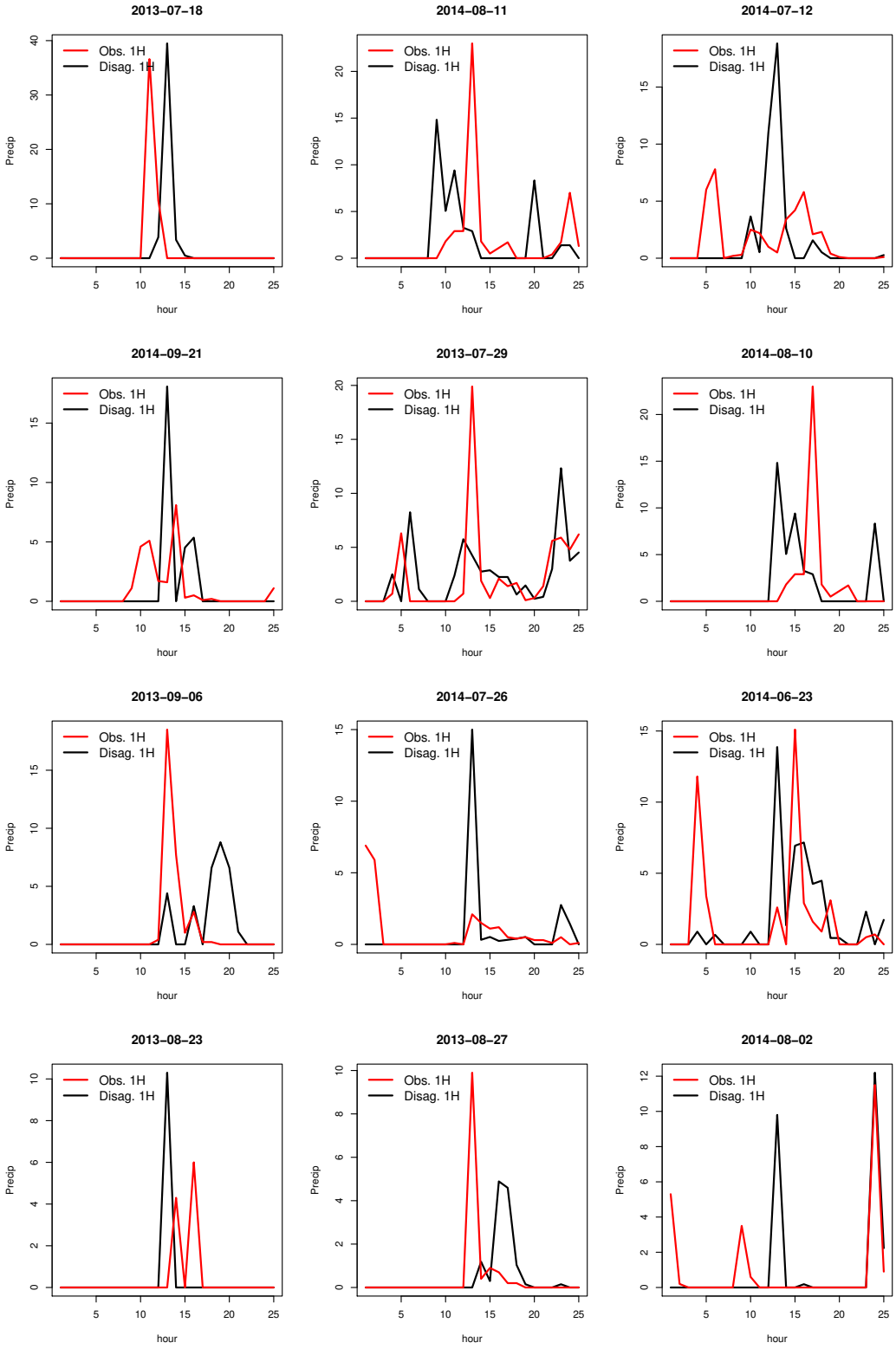


Figure 6: Examples of different sub-daily temporal structures of precipitation during the period 1990-2014, at the stations 'THU' (Obs. 1H) and 'BER' (Disag. 1H).

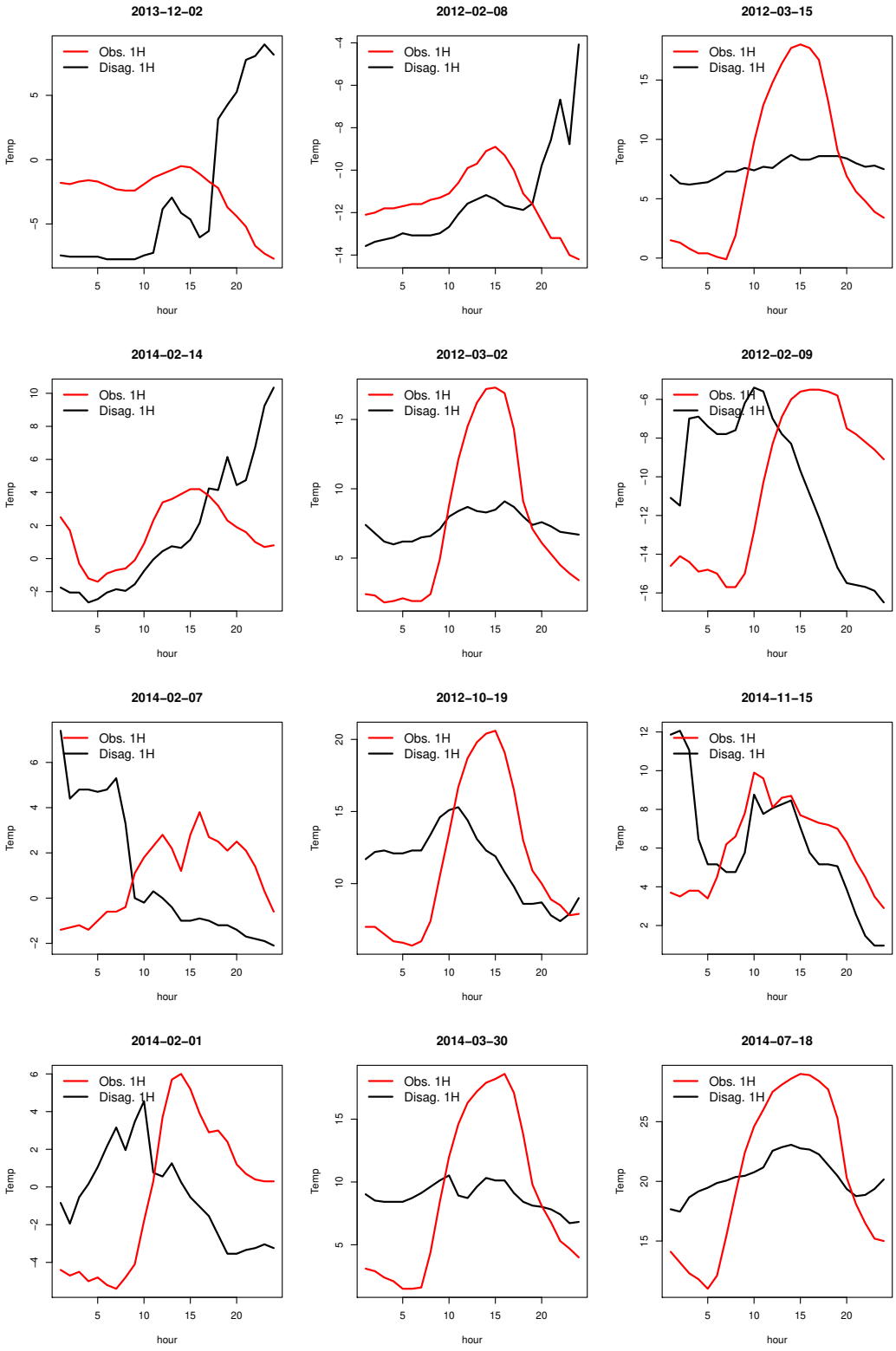


Figure 7: Examples of different sub-daily temporal structures of temperature during the period 1990-2014, at the station 'LAG' (Obs. 1H) and 'NAP' (Disag. 1H).

4.3 Comparison between areal observations and areal values obtained from disaggregated observations

For each sub-catchment, areal estimates are computed with the Thiessen Polygon method. Here we denoted areal precipitation and temperature values at an hourly scale, obtained from the disaggregated observations, by $\tilde{\mathcal{P}}_h^1$ and $\tilde{\mathcal{T}}_h^1$, respectively.
For the sake of comparison, we also compute areal estimates directly from observed data, using all the available stations during the period 2009-2013. As can be seen in Figures 8 and 9, the gauge networks are dense during this period, although the number of stations in operation vary. In comparison, the density of daily gauges where the disaggregated observations are computed is not as dense, especially for temperature.
Figures 10 and 11 compare these two products:

- areal estimates directly computed from hourly observations $\mathcal{P}_h^1$ and $\mathcal{T}_h^1$,
- areal estimates computed from daily observations disaggregated at an hourly scale $\tilde{\mathcal{P}}_h^1$ and $\tilde{\mathcal{T}}_h^1$.

Both products are computed from a limited number of stations and the differences between the two products mainly indicate the difference between the daily and hourly gauge networks. Figure 10 shows that the 90th percentile of absolute deviations between $\mathcal{P}_h^1$ and $\tilde{\mathcal{P}}_h^1$ do not exceed 0.2mm/h for most of the catchments. For a few catchments, either daily or hourly gauges are distant and explain these discrepancies. For temperature, Figure 10 shows that the 90th percentile of absolute deviations between $\mathcal{T}_h^1$ and $\tilde{\mathcal{T}}_h^1$ are below 4 degrees for most of the catchments.

5 Refinements of the disaggregation method

The disaggregation method presented is a first benchmark model and several options could be investigated to improve this version. For temperature, the selection of the closest hourly gauge could be modified, for example by penalizing strongly difference of elevations between the stations. The inclusion of other predictors such as the daily temperature field or a pre-selection of the analog dates based on a seasonal filter or an atmospheric circulation models could also be considered.

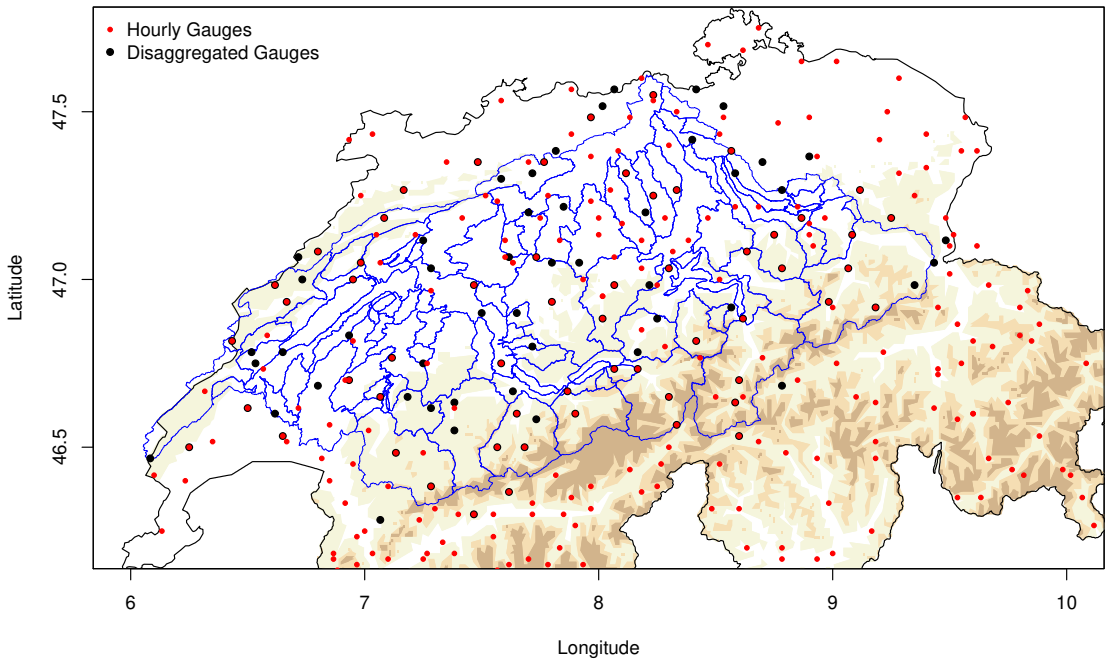


Figure 8: Complete network of hourly gauges for precipitation data (red dots) to be compared to the daily gauges where observations have been disaggregated (black dots).

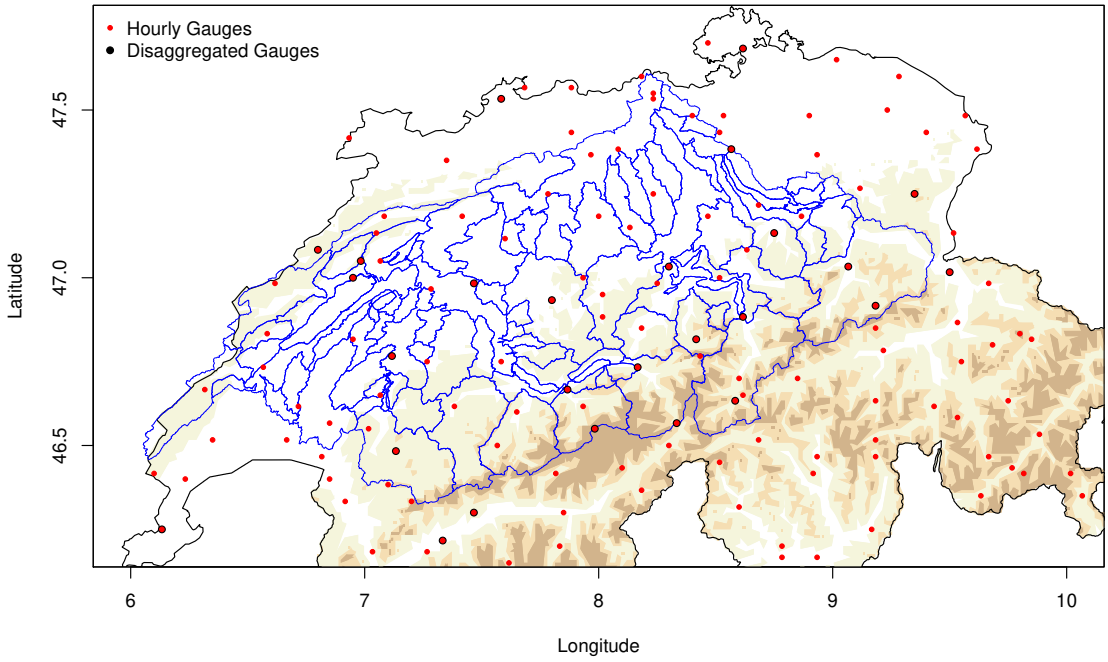


Figure 9: Complete network of hourly gauges for temperature data (red dots) to be compared to the daily gauges where observations have been disaggregated (black dots).

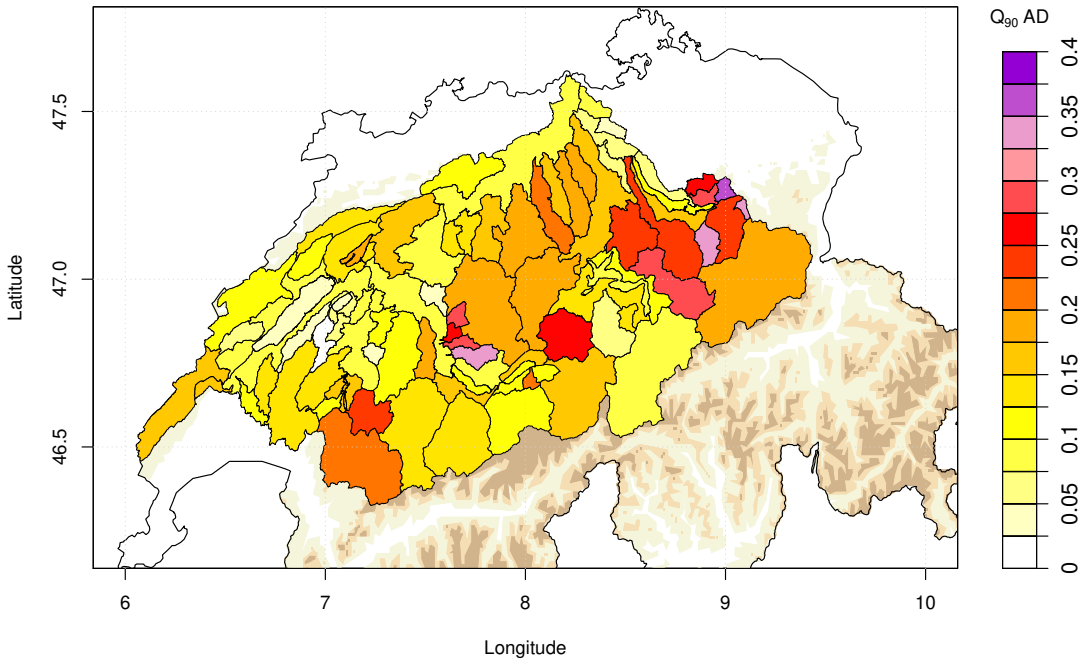


Figure 10: 90th percentile of absolute deviations between hourly areal precipitation observations computed from the daily gauges (disaggregated at an hourly scale) and the hourly gauges.

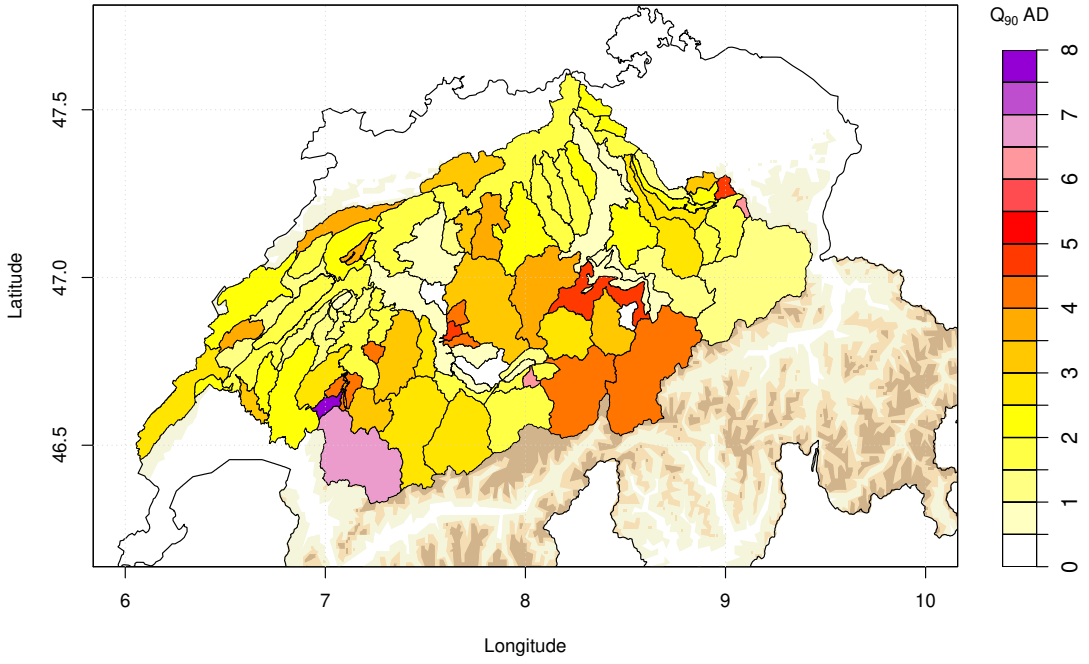


Figure 11: 90th percentile of absolute deviations between hourly areal temperature observations computed from the daily gauges (disaggregated at an hourly scale) and the hourly gauges.

10.3 GWEX weather generator

10.3.1 Stochastic generation of multi-site daily precipitation focusing on extreme events

Stochastic generation of multi-site daily precipitation focusing on extreme events

Guillaume Evin[1], Anne-Catherine Favre[1], and Benoit Hingray[1]

[1]Univ. Grenoble Alpes, CNRS, IRD, Grenoble INP*, IGE, F-38000 Grenoble, France

Correspondence to: Guillaume Evin (guillaume.evin@irstea.fr)

Abstract. Many multi-site stochastic models have been proposed for the generation of daily precipitation, but they generally focus on the reproduction of low to high precipitation amounts at the concerned stations. This paper proposes significant extensions to the multi-site daily precipitation model introduced by Wilks in the aim of reproducing the statistical features of extremely rare events (in terms of frequency and magnitude) at different temporal and spatial scales. In particular, the first extended version integrates heavy-tailed distributions, spatial tail dependence, and temporal dependence in order to obtain a robust and appropriate representation of the most extreme precipitation fields. A second version enhances the first version using a disaggregation method. The performance of these models is compared at different temporal and spatial scales on a large region covering approximately half of Switzerland. While daily extremes are adequately reproduced at the stations by all models, including the benchmark Wilks version, extreme precipitation amounts at larger temporal scales (e.g. 3-day amounts) are clearly underestimated when temporal dependence is ignored.

1 Introduction

Stochastic precipitation generators are often employed in risk assessment studies to estimate the return periods of very rare flooding events (e.g. 10,000-year events). The observed series of streamflows are too short to produce reliable estimations of very rare and large floods. Typically, extreme hydrological events can be reproduced using long series of simulated precipitation data as input to hydrological models (Lamb et al., 2016).

In the last two decades, a number of precipitation models have been proposed to deal with the temporal and spatial properties of daily precipitation, for both intermittency and amount, and all have different strengths and weaknesses. Many of these models use exogenous variables to predict the statistical properties of precipitation using generalized linear models (Chandler and Wheater, 2002; Mezghani and Hingray, 2009; Serinaldi and Kilsby, 2014b), atmospheric analogs (Lafaysse et al., 2014), or modified Markov models (Mehrotra and Sharma, 2010). Introducing a link between exogenous atmospheric variables can be used to reconstruct past events, make predictions, or downscale GCM-based simulations of future climate. Such models are classically referred to as statistical downscaling models (see Maraun et al., 2010, for a review). Closely related to this approach, weather 'types' or 'regimes' (Ailliot et al., 2015) can be used to specifically account for different atmospheric

*Institute of Engineering Univ. Grenoble Alpes

circulation patterns. Using Hidden Markov Model (HMM) with transitions between these weather states, stochastic weather generators can then simulate various aspects of the precipitation process (Rayner et al., 2016).

Alternatively, purely stochastic precipitation models can be used. These can be broadly classified into three main types:

- **Resampling methods:** The stochastic generation of precipitation fields can be performed using resampling techniques such as the K-nearest neighbors (Buishand, 1991; Yates et al., 2003). Unobserved precipitation amounts can be obtained using perturbation techniques (Sharif and Burn, 2007).

- **Random fields:** Spatio-temporal precipitation models can simulate precipitation fields over a regular grid. This approach is particularly useful for hydrological applications, since areal precipitation values over a basin are obtained directly. Poisson cluster-based models (Burton et al., 2008, 2010; Leonard et al., 2008; McRobie et al., 2013) randomly simulate rain disk cells, with random centers, radius and intensity, over the study area. Meta-Gaussian models (Vischel et al., 2009; Kleiber et al., 2012; Allard and Bourotte, 2015; Baxevani and Lennartsson, 2015; Bennett et al., 2017) are based on truncated and transformed random Gaussian fields. Closely related, the turning band method can be used to simulate intermittent precipitation fields with different type of advection (Leblois and Creutin, 2013). These model structures are appealing since they are able to simulate realistic precipitation fields at fine spatial scales. However, their complexity leads to numerous technical issues during parameter estimation and simulation, notably in terms of computational cost. Moreover, they are usually unable to represent large regions comprising very distinct precipitation regimes.

- **Statistical multi-site models:** In this last type of weather generator, the properties of precipitation are directly fitted at a limited number of stations using different statistical structures. This type of generator preserves the inter-dependency between all pairs of stations, even when the area under study exhibits different precipitation regimes. Bárdossy and Pegram (2009) and Rasmussen (2013) combine a multivariate autoregressive process and transformations (V-transform, power transformation) to simultaneously model precipitation occurrence and amount. More precisely, with these models, transformed precipitation amounts follow truncated distributions. Alternatively, Wilks (1998) proposes a multi-site model in which precipitation occurrence and amount are handled separately. Several extensions to this popular structure have been proposed in the literature. Thompson et al. (2007) reformulate the Wilks model as a hidden Markov model, inferring three precipitation states ('dry', 'light' and 'heavy'). Mehrotra and Sharma (2007b) apply semi-parametric techniques to add more flexibility to the spatial structure of precipitation occurrence and amount. Srikanthan and Pegram (2009) propose a modified version in which daily, monthly and annual amounts are nested such that precipitation statistics are preserved for all these levels of aggregation.

Mehrotra et al. (2006) compare three different precipitation models, the Wilks model, a HMM and a resampling approach, and provide strong arguments in favor of the Wilks model in terms of performance, computation time, model, and level of complexity of the model structure. Furthermore, as indicated above, this model offers a flexible structure which can be applied to a large number of stations with very different precipitation regimes (like in mountainous areas). This paper presents several significant extensions of the Wilks precipitation model, referred to as GWEX versions, which will be used to generate long scenarios. These extensions aim at fitting the most extreme precipitation amounts at different temporal (1-day and 3-day amounts)

and spatial scales. Novel components are thus introduced in GWEX, including robust estimation methods (regionalization methods) for critical parameters impacting directly on the behavior of extreme precipitation at each station. Also included are recent advances in the choice of the marginal distributions for daily precipitation amounts. Using 15,029 long daily precipitation records ($>$ 50 years) from around the world, Papalexiou et al. (2013) conclude that heavy-tailed distributions are generally in better agreement with observed precipitation extremes. Follow-up studies (Papalexiou and Koutsoyiannis, 2013; Serinaldi and Kilsby, 2014a) apply extreme value theory to annual maxima and "peaks over threshold" (POTs) of a large subset of these records and confirm that extreme daily precipitation is not adequately represented by light-tailed distributions. Based on statistical tests on 90,000 station records of daily precipitation, Cavanaugh et al. (2015) also come to the same conclusion. These findings have important implications for precipitation models:

- Light-tailed distributions such as exponential, Gamma, and Weibull distributions, which are applied in the vast majority of the existing precipitation models, often lead to an underestimation of extreme daily precipitation amounts.
- While non-parametric densities with Gaussian kernels (Mehrotra and Sharma, 2007a, 2010) offer the flexibility to fit the observed range of precipitation amounts, their tail also belongs to the domain of attraction of the Gumbel distribution and suffers from the same drawbacks.

Alternatively, current statistical procedures consisting in fitting a flexible distribution to the bulk of the observations and using it for extrapolation are highly questionable, as major assumptions are usually violated (Klemeš, 2000a, b). Since the tail of the distribution on precipitation amounts at each station will dictate the generation of the most extreme precipitation events, important features of GWEX are:

- application of a heavy-tailed distribution to precipitation amounts at each station (Naveau et al., 2016),
- determination of robust estimates of the shape parameter of this distribution, which indicates the heaviness of the tail, using a regionalization approach, as in Evin et al. (2016).

Furthermore, following Bárdossy and Pegram (2009), GWEX also employs the copula theory to introduce a tail dependence between the precipitation amounts simulated at the different stations. The second version of the GWEX model includes a disaggregation method, the observed precipitation amounts being fitted at a 3-day scale in a first step. This paper compares the performance of the different model versions and assesses the impact of the different statistical components (e.g. heavy-tailed distribution, tail dependence, etc.).

We first describe the study area in Section 2. The features of different multi-site precipitation models are then described in Section 3. The evaluation framework, presented in Section 4, aims at assessing the performance of these models at different spatial and temporal scales. Section 5 presents an application of these daily precipitation models to 105 stations located in Switzerland, with a summary of the results focusing on the reproduction of extreme events. Finally Section 6 presents our conclusions.

2 Data and study area

The Aare River basin covers the northern part of the Swiss Alps and has an area of 17,700 km^2. Basin elevations approximately range from 310 m.a.s.l. in Koblenz (entrance to Germany in the north) to 4270 m.a.s.l. at the Finsteraarhorn summit (in the south of the basin). The mean annual precipitation for the basin as a whole is 1300 mm. The basin can be divided into five main sub-basins with different hydrometeorological regimes highly governed by regional terrain features (Jura mountains in the north-west; Northern Alps in the south of the basin and lowlands in the middle).

Figure 1 shows the location of the 105 precipitation stations used for the development and evaluation of weather generators. Located within or close to the Aare River basin, they correspond to the stations for which long daily time series of observations with less than 3 years of missing data are available over the period 1930-2014. The 105 precipitation stations cover relatively well the Aare River basin.

The proposed precipitation models are designed to simulate, via a conceptual hydrological model, flood scenarios for the whole Aare River basin and for its different sub-basins. For Switzerland, Froidevaux et al. (2015) show that the generation of floods is mainly influenced by areal precipitation amounts accumulated over short periods (e.g. 1 to 3 days). These results are obtained by analyzing a wide variety of basins, their areas ranging from 10 km^2 to 12,000 km^2. Therefore, the properties of the weather scenarios must be evaluated at different spatial and temporal scales, from the high resolutions required to simulate the hydrological behavior of the system (e.g. sub-daily, 100 km^2) to lower resolutions relevant at the scale of the entire basin (e.g. n-days, 17,700 km^2). In this study, the performance of the different precipitation models is evaluated at the station scale, at the scale of 15 and 5 sub-basins partitioning the Aare River basin, and at the scale of the entire study area (see Section 5). Note that for those evaluations, areal estimates of precipitation are obtained from the precipitation amounts at the stations using the Thiessen polygon method.

3 Multi-site precipitation model

As indicated above, GWEX refers to multi-site precipitation models that rely strongly on the structure proposed by Wilks (1998). At each location k, let $P_t(k)$ be a random variable representing the accumulated precipitation over day t. The structure proposed by Wilks considers a hidden occurrence process $X_t(k)$ that can be represented by a two-state Markov chain as follows:

$$X_t(k) = \begin{cases} 0, & \text{if day } t \text{ is dry at location } k. \\ 1, & \text{if day } t \text{ is wet at location } k. \end{cases} \tag{1}$$

Precipitation amount $P_t(k)$ is then defined as:

$$P_t(k) = Y_t(k) X_t(k) \tag{2}$$

where $Y_t(k)$ is a random variable describing the non-zero precipitation amounts. Non-zero precipitation amounts $Y_t(k)$ are thus modeled independently of precipitation occurrences $X_t(k)$, which act as a mask.

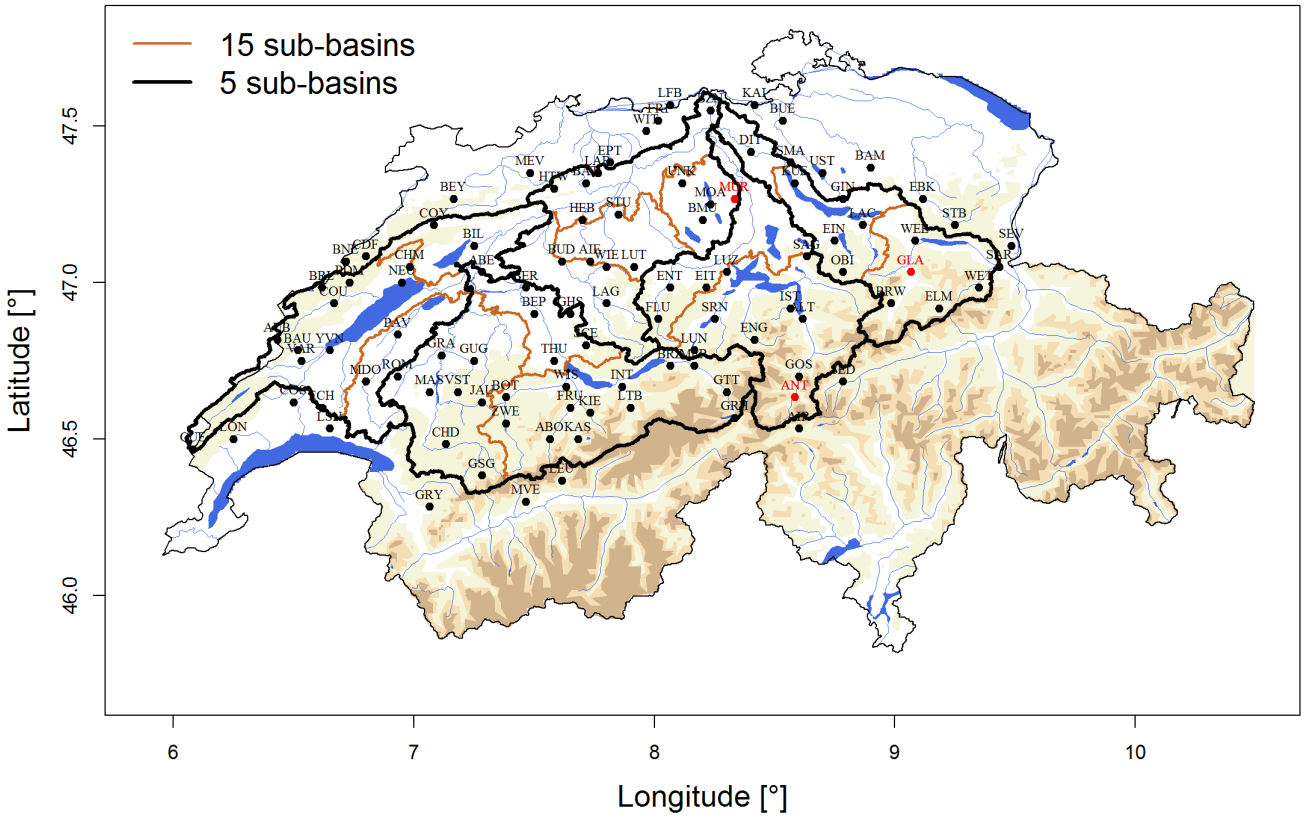


Figure 1. Location of the 105 precipitation stations in Switzerland. Different partitions of the Aare River basin into 5 and 15 sub-basins are shown.

3.1 Precipitation occurrence process

3.1.1 At-site occurrence process

At each location, the temporal persistence of dry and wet events is introduced with a p-order Markov chain model for $X_t(k)$ so that:

$$\Pr\{X_t(k) = 1 | X_{t-1}(k), \ldots, X_1(k)\} = \Pr\{X_t(k) = 1 | X_{t-1}(k), \ldots, X_{t-p}(k)\}, \quad (3)$$

i.e. the probability of having a wet day at time t depends only on the p previous states, for days $t-1, \ldots, t-p$. While many authors suppose that a first-order Markov is sufficient (e.g. Wilks, 1998; Keller et al., 2015), Srikanthan and Pegram (2009) apply a 4-order Markov chain and show that it improves the reproduction of dry/wet period lengths. In this study, different orders for this Markov chain are considered.

At each site, the probability of having a wet day at day t is given by the transition probability $\Pr\{X_t(k) = 1 | X_{t-1}(k) = i_1, \ldots, X_{t-p}(k) = i_p\}$, where $i_1, \ldots, i_p$ are equal to 0 or 1. This Markov chain is thus fully characterized by a transition matrix Π with dimension 2^p.

3.1.2 Spatial occurrence process

The spatial dependence of the precipitation states $X_t(k)$ is modeled using an unobserved Gaussian stochastic process $\mathbf{U}_t = \{U_t(1), \ldots, U_t(K)\}$, where K is the number of stations. Here, Gaussian random variables $U_t(k), k = 1, \ldots, K$, are temporally independent and $\mathbf{U}_t$ follows a multivariate normal distribution:

$$\mathbf{U}_t \sim N(0, \Omega_X), \tag{4}$$

where $\Omega_X = \{\omega_{kl}\}$ is a positive-definite correlation matrix. At any location k, the precipitation state $X_t(k)$ is assumed to be completely determined by $U_t(k)$ and the previous p states at the same location. Specifically, if $X_{t-1}(k) = i_1, \ldots, X_{t-p}(k) = i_p$, and $p_1 = \Pr\{X_t(k) = 1 | X_{t-1}(k) = i_1, \ldots, X_{t-p}(k) = i_p\}$, then

$$X_t(k) = \begin{cases} 1, & \text{if } U_t(k) \leq \Phi^{-1}(p_1). \\ 0, & \text{otherwise,} \end{cases} \tag{5}$$

where $\Phi[.]$ indicates the standard Gaussian cumulative distribution function.

Let $\rho_{kl} = \text{Corr}(X_t(k), X_t(l))$ denote the inter-site correlation between the states $X_t(k)$ and $X_t(l)$. Following Srikanthan and Pegram (2009), ρ_{kl} can be expressed as:

$$\rho_{kl} = \frac{\pi_{00}(k,l) - \pi_0(k)\pi_0(l)}{\sqrt{\pi_0(k)\pi_1(k)}\sqrt{\pi_0(l)\pi_1(l)}}, \tag{6}$$

where $\pi_0(s) = \Pr\{X_t(s) = 0\}$ and $\pi_1(s) = \Pr\{X_t(s) = 1\}$ denote the probabilities of having dry and wet states at location s, respectively, and $\pi_{00}(k,l) = \Pr\{X_t(k) = 0, X_t(l) = 0\}$ denotes the joint probability of having dry states at both locations k and l.

The relationship between ω_{kl} and ρ_{kl} is not direct since the temporal persistence of dry and wet events introduced at each station with a Markov chain also influences ρ_{kl} (Wilks, 1998). Figure 2 illustrates this relationship, obtained for the month of July via Monte-Carlo simulations, for two close stations, GOS and ANT. In a first step, transition probabilities with a Markov chain of order 4 are estimated for these two stations. Given these transition probabilities, stochastic simulations of occurrence are then generated for different values of ω_{kl}, leading to different values of ρ_{kl}. Since this relationship is monotonic (see Fig. 2), it can be used to identify the value ω_{kl} leading to a specific $\hat{\rho}_{kl}$, namely the empirical value obtained from the observed time series of occurrence. The estimate of ω_{kl} is found by iterating until the evaluation of the correlation between the simulated precipitation states, ρ_{kl}, matches $\hat{\rho}_{kl}$. Note that a very high value for $\hat{\rho}_{kl}$ cannot always be reached, even if $\omega_{kl} = 1$. This is however a situation which rarely occurs in practice.

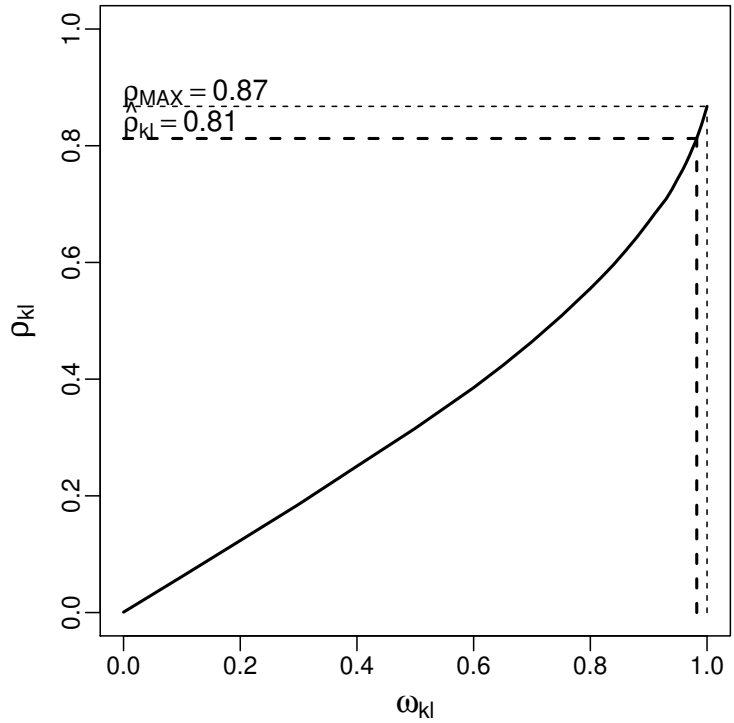


Figure 2. Illustration of the relationship between ω_{kl} and ρ_{kl} for the month of July and for stations GOS and ANT. A Markov chain of order 4 is considered in this example. The correlation between the observed states is $\hat{\rho}_{kl} = 0.81$ and can be reproduced using a bivariate Gaussian distribution with a correlation parameter of $\omega_{kl} = 0.98$. The maximum correlation ρ which can be obtained if $\omega_{kl} = 1$ is $\rho_{MAX} = 0.87$.

3.2 Precipitation intensity process

Given the occurrence of precipitation $X_t(k)$ at different locations k, GWEX models generate the amounts of precipitation $Y_t(k)$ using:

- marginal heavy-tailed distributions,
- a tail-dependent spatial distribution,
- an autocorrelated temporal process.

3.2.1 Marginal distributions

At a given location k, daily precipitation has often been modeled by light-tailed distributions: exponential and Weibull distributions (Bárdossy and Pegram, 2009); gamma distributions (Srikanthan and Pegram, 2009; Mezghani and Hingray, 2009); mixture of exponential distributions (Wilks, 1998; Keller et al., 2015); mixture of gamma distributions (Chen et al., 2014).

However, as shown by many recent studies on a very large number of daily precipitation series (Papalexiou et al., 2013; Serinaldi and Kilsby, 2014a; Cavanaugh et al., 2015), exponentially decaying tails often result in a severe underestimation of extreme event probabilities. The introduction of a heavy-tailed distribution is thus crucial for the reproduction of the most extreme precipitation events (Hundecha et al., 2009).

In this work, the distribution representing the precipitation intensity at each location, $Y_t(k)$, is the E-GPD distribution. This distribution was first proposed by Papastathopoulos and Tawn (2013) who referred to it as an extended GP-Type III distribution and has since been shown to adequately model the whole range of precipitation intensities (Naveau et al., 2016). Compared to other heavy-tailed distributions applied to daily precipitation amounts (e.g. mixtures of GPD and gamma distribution, see Vrac and Naveau, 2007), the E-GPD is parsimonious and provides a very good compromise between flexibility and stability, which is an essential feature for extrapolation.

This distribution can be described by a smooth transition between a gamma-like distribution and a heavy-tailed Generalized Pareto distribution (GPD). This transition is obtained via a transformation function, $G(\nu)$, such that the whole range of precipitation intensities is modeled without a threshold selection (Naveau et al., 2016):

$$F_Y\{Y_t(k)\} = G\Big[H_\xi\big\{Y_t(k)/\sigma\big\}\Big], \tag{7}$$

where

$$H_\xi(z) = \begin{cases} 1-(1+\xi z)_+^{-1/\xi} & \text{if } \xi \neq 0, \\ 1-e^{-z} & \text{if } \xi = 0, \end{cases} \tag{8}$$

with $a_+ = \max(a,0)$, is the standard cumulative distribution function of the GPD, $\sigma > 0$ is a scale parameter and $G(\nu) = \nu^\kappa, \kappa > 0$. Thus, a 3-parameter set $\{\sigma, \kappa, \xi\}$ needs to be estimated at each station.

3.2.2 Spatial and temporal dependence of precipitation amounts

Spatial and temporal dependence of precipitation amounts is represented using a Multivariate Autoregressive model of order 1 (MAR(1)). A MAR(1) process has been used by different authors (Bárdossy and Pegram, 2009; Rasmussen, 2013) to simultaneously represent spatial and temporal dependences. Let $\mathbf{Z}_t$ denote a vector of K Gaussian random variables with mean 0 defined as:

$$Z_t(k) = \Phi^{-1}\big[F_Y\{Y_t(k)\}\big]. \tag{9}$$

The stochastic Gaussian process $\mathbf{Z}_t$ is assumed to follow a MAR(1) process defined as follows:

$$\mathbf{Z}_t = \mathbf{A}\mathbf{Z}_{t-1} + \epsilon_t, \tag{10}$$

where $\mathbf{A}$ is a $K \times K$ matrix and ϵ_t is an innovation term described by a random $K \times 1$ noise vector. The elements of ϵ_t have zero means and are independent of the elements of $\mathbf{Z}_{t-1}$. The covariance matrix of ϵ_t is denoted by Ω_Z. Following Bárdossy

and Pegram (2009), $\mathbf{A}$ is taken to be a diagonal matrix with diagonal elements that are the lag-1 serial correlation coefficients of the intensity process $Y_t(k)$. The matrix Ω_Z can be expressed as:

$$\Omega_Z = \mathbf{M}_0 - \mathbf{A}\mathbf{M}_0'\mathbf{A}, \tag{11}$$

where $\mathbf{M}_0$ is the covariance matrix of $\mathbf{Z}_t$, which indicates the degree of spatial dependence between each pair of stations, and $\mathbf{M}_0'$ is its transpose.

Innovations ϵ_t are often assumed to follow a standard multivariate normal distribution. However, the upper tail dependence of the multivariate normal distribution is 0, which means that extreme precipitation amounts simulated at the different sites are not spatially dependent. To introduce a tail dependence between at-site extremes, a possibility is to use a Student copula to represent the dependence structure of ϵ_t, providing an additional parameter, ν, related to the tail dependence. Both dependence structures will be considered in the following.

3.3 Parameter estimation

3.3.1 Occurrence process

Following Wilks (1998), parameters related to the occurrence process $X_t(k)$ are estimated using the method of moments, i.e. using the empirical counterparts of the parameters. Observed states are first obtained using a low precipitation threshold (e.g. 0.2 mm). The matrix Π of transition probabilities are then estimated directly by the proportion of wet days $X_t(k) = 1$ following observed sequences $\{X_{t-1}(k), \ldots, X_{t-p}(k)\}$. Concerning the spatial occurrence process, $\hat{\rho}_{kl}$ estimates are obtained using the empirical counterparts of π_{00}, π_0 and π_1 (see Eq. 6), which correspond respectively to the proportion of days for which dry states are observed simultaneously at two locations ($\hat{\pi}_{00}$) and to the proportions of dry days $\hat{\pi}_0$ and wet days $\hat{\pi}_1$. The correlation matrix $\hat{\Omega}_X$ is then composed of the cross-correlations $\hat{\omega}_{kl}$ obtained for all possible pairs of stations. If $\hat{\Omega}_X$ is not positive-definite, the closest positive-definite matrix is considered (Rousseeuw and Molenberghs, 1993; Rebonato and Jaeckel, 2011). Furthermore, the seasonality of the occurrence process is taken into account by estimating these parameters on a monthly basis.

3.3.2 Intensity process

E-GPD distributions are first fitted to precipitation amounts available at each location k. Local estimations of the GPD tail exhibiting a lack of robustness, we propose to estimate the ξ parameter of the E-GPD (see Eq. 8) using a regionalization method similar to that of Evin et al. (2016), which can be summarized as follows:

1. Following Burn (1990), for each station, a region-of-influence (RoI) is delimited by a circle around the site, the radius being determined using homogeneity tests. All the stations inside this RoI are then considered homogeneous up to a scale factor.

2. The ξ parameters are then estimated with the maximum likelihood method using the precipitation observations from all the stations inside the RoI.

This regionalization method is applied to the precipitation data available from 666 stations in Switzerland, for 4 different seasons:

- **Winter:** December, January and February,
- **Spring:** March, April and May,
- **Summer:** June, July and August,
- **Autumn:** September, October and November.

In this work, the estimation of the ξ parameter is bounded below by 0. When $\xi < 0$, the E-GPD distribution has an upper bound. As shown by many recent studies (e.g. Serinaldi and Kilsby, 2014a), negative estimates of ξ are usually due to parameter uncertainty and are not realistic. The two remaining parameters of the E-GPD, the scale parameter σ and the parameter of the transformation κ, are estimated from the observations available at that station. Here, we use a method of moments based on probability weighted moments (see Naveau et al., 2016, for further details).

Concerning the spatial and temporal dependence of precipitation amounts, direct estimates of $\mathbf{M}_0$ and $\mathbf{A}$ cannot be obtained since non-zero precipitation amounts $Y_t(k)$ are not observed. Here, we follow the methodology proposed by Wilks (1998) and Keller et al. (2015). For each pair of stations, we generate long sequences of precipitation amounts $P_t(k)$ using the estimated parameters of the occurrence process ($\hat{\Pi}$ and $\hat{\omega}_{kl}$), the parameters of the marginal distributions and a correlation coefficient $m_0(k,l)$ indicating the degree of spatial dependence. Similarly to the occurrence process, $\hat{m}_0(k,l)$ is then found iteratively by matching the correlation between these long random streams with the observed correlation $\text{Corr}(P_t(k), P_t(l))$ (see Wilks, 1998; Keller et al., 2015, for further details). The correlation matrix $\hat{\mathbf{M}}_0$ is then composed of the cross-correlations $\hat{m}_0(k,l)$ obtained for all possible pairs of stations. For each station, the estimates of the lag-1 serial correlation coefficients of the matrix $\mathbf{A}$ are obtained using the same simulation approach.

The matrix $\hat{\Omega}_Z$, i.e. the estimate of the covariance matrix of the innovations ϵ_t, is then obtained using Eq. 11. Since $\hat{\Omega}_Z$ is not necessarily positive-definite (see Eq. 11), the closest positive-definite matrix is taken as the covariance matrix of ϵ_t if necessary. Given $\hat{\Omega}_Z$, the parameter ν is estimated by maximizing the likelihood, as described in McNeil et al. (2005, Section 5.5.3.).

Similarly to the occurrence process, the seasonal aspect of the precipitation intensity is taken into account by performing the parameter estimation for each month, on a 3-month moving window.

3.4 Model versions

Different versions of the proposed multi-site precipitation model are considered in this paper, each corresponding to different extensions of the Wilks model. A flowchart summarizing the increasing complexity of these models is presented in Figure 3.

3.4.1 Wilks

A first benchmark version of the multi-site model, referred to here as 'Wilks', is considered. It closely matches the multi-site model proposed by Wilks (1998). In particular:

- The at-site occurrence process is a Markov chain of order 1.
- The marginal distribution on precipitation amounts is a mixture of exponential distribution, for which the pdf is defined as:

$$f(x) = \frac{w}{\beta_1} \exp\Big(-\frac{x}{\beta_1}\Big) + \frac{1-w}{\beta_2} \exp\Big(-\frac{x}{\beta_2}\Big). \tag{12}$$

  The parameters w, β_1 and β_2 are estimated using the Expectation-Maximization (EM) method (Dempster et al., 1977).

- Precipitation amounts are not considered to be temporally correlated, i.e. the matrix $\mathbf{A}$ in Eq. (10) is a zero matrix. Furthermore, innovations ϵ_t follow a standard multivariate normal distribution and represent the spatial correlations.

3.4.2 Wilks_EGPD

A modified Wilks version is considered, for which the at-site occurrence process is a Markov chain of order 4 and the mixture of exponential distributions is replaced by the E-GPD distribution. As indicated above, Srikanthan and Pegram (2009) show that a 4-order Markov chain improves the reproduction of dry/wet period lengths. This direct extension of the Wilks model is used to illustrate the impact of using a Markov chain of order 4 compared to order 1. Differences in performance between a heavy-tailed distribution (E-GPD) and a low-tailed distribution (mixture of exponentials) will be highlighted.

3.4.3 GWEX

The initial GWEX model has the following characteristics:

- The at-site occurrence process is a Markov chain of order 4.
- The marginal distribution for precipitation amounts is the E-GPD distribution.
- Precipitation amounts follow a MAR(1) process with innovations modeled by a Student copula.

3.4.4 GWEX_Disag

In this paper, an alternative version, referred to as GWEX_Disag, is also proposed. GWEX_Disag is applied to 3-day precipitation amounts and has the same characteristics as GWEX, except that:

- The at-site occurrence process is a Markov chain of order 1.
- A threshold of 0.5 mm separates dry and wet states.

With GWEX_Disag, daily scenarios are first generated at a 3-day scale and then disaggregated at a daily scale using a method of fragments (e.g. Wójcik and Buishand, 2003). Simulated 3-day amounts are disaggregated using the temporal structures of the closest observed 3-day amounts, in terms of similarity of the spatial fields. The same observed 3-day sequence is thus used to disaggregate the 3-day amounts simulated at the 105 stations, which ensures the spatial coherence of these disaggregated amounts. Details of the disaggregation method are provided in Appendix A. Compared to GWEX, GWEX_Disag offers the following advantages:

- 3-day precipitation amounts are directly modeled and have a better chance to be adequately reproduced,
- the disaggregation of 3-day precipitation amounts creates an inherent link between the occurrence and the intensity processes. For very extreme precipitation events, we can expect these processes to be dependent (higher chance to be in a wet state over the whole Aare River basin, as well as large and persistent precipitation amounts).

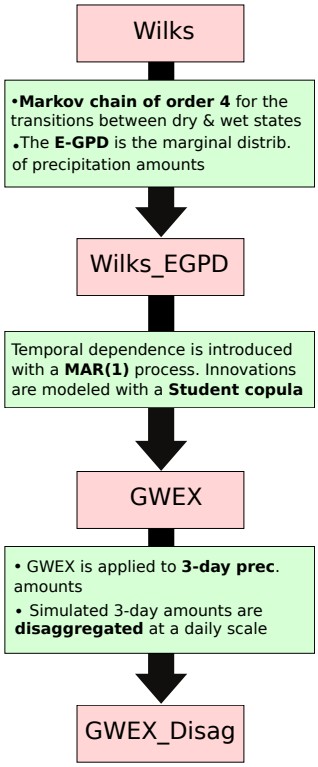


Figure 3. Flowchart of the different model versions. The differences between the models are summarized inside green boxes.

4 Multi-scale evaluation

The proposed stochastic models intend to preserve the most critical properties of precipitation at different spatial and temporal scales, especially extreme precipitation amounts. For hydrological applications, it can be assumed that a precipitation model preserving these properties has a better chance of adequately reproducing flood properties for small sub-basins as well as for large basins. This statement is supported by empirical evidence provided by Froidevaux (2014) and Froidevaux et al. (2015) for our study area (i.e. Switzerland). Using 60 years of gridded precipitation data, Froidevaux et al. (2015) show that, in Switzerland, high discharge events are usually triggered by meteorological events with a duration of several days, in late summer and autumn. Typically, the 2-day precipitation sum before floods is most correlated with flood frequency and flood magnitude.

The performance of the different multi-site precipitation models is thus assessed for multiple spatial and temporal scales. We investigate whether or not the statistical properties of precipitation data are adequately reproduced at the scale of the stations and for different partitions of the Aare River basin (see Figure 1). In order to achieve this, 100 daily precipitation scenarios are generated, each scenario having a length of 100 years.

For the different evaluated statistics, performance is categorized according to the comprehensive and systematic evaluation (CASE) framework proposed by Bennett et al. (2017). The CASE framework enables a systematic comparison of stochastic models and offers a consistent way of computing the performance metrics, which is important in order to obtain a fair assessment of the strengths/weaknesses of the different model versions. This approach consists in assigning one of three categories: 'good', 'fair' and 'poor' performance, to each metric, according to the agreement between the observed metric and the simulated metrics computed from the 100 scenarios. Table 1 summarizes the tests leading to each performance category. 'Good' performance is obtained when the observed metric is inside the 90% probability limits of the 100 simulated metrics (case 1). It indicates that simulated metrics are in good agreement with the observed metric. However, an observed metric can obviously lie outside these limits without necessarily indicating a failure of the model. In this case, 'fair' performance may assigned if either of the following two rules is satisfied:

1. Case 2: The observed metric is outside the 90% probability limits but within three standard deviations of the simulated mean, which corresponds to the 99.7% probability limits if we assume that the uncertainty in the statistics is normally distributed. This case covers the situation where we could expect that the observed metric is outside the 90% limits due to sampling uncertainty.

2. Case 3: The absolute relative difference $|(S_{obs} - \bar{\mathbf{S}}_{sim})/S_{obs}|$ between the observed metric S_{obs} and the mean of the simulated metrics $\bar{\mathbf{S}}_{sim}$ is 5% or less. If the variability of the simulated metrics is very small, it can happen that the observed metric lie outside the 99.7% limits without being too far from the simulated mean in terms of relative difference.

Otherwise, we consider that performance is 'poor', indicating that the model fails to reproduce this particular statistical property.

In summary, 'good' performance represents cases for which the observed metric is clearly well reproduced by the model, whereas 'fair' performance indicates a reasonable match between the observed and the simulated metrics. The number of metrics for which 'poor' performance is obtained is thus the first criteria indicating the overall performance of a model.

Table 1. Performance categorization criteria from Bennett et al. (2017).

Performance Classification	Key	Test
'good'		Observed metric inside 90% limits (case 1)
'fair'		Observed metric outside 90% limits but within the 99.7% limits (case 2) OR absolute relative difference between the observed metric and the average simulated metrics is 5% or less (case 3)
'poor'		Otherwise (case 4)

For illustration purposes, we also present the results of the evaluation for three precipitation stations corresponding to different hydrological regimes (see Table 2). Figure 1 shows the 3 (out of 105) selected precipitation stations. Station ANT (at Andermatt) is located in a glacial basin, station GLA (at Glarus) in a nival basin and station MUR (at Muri) in a pluvial basin.

Table 2. Hydrological regimes and characteristics of extreme floods in Switzerland (Froidevaux, 2014).

	Mean elevation [m]	Season	Triggering events
Glacial	> 1900	summer	showers + snow melt
Nival	$1200 - 1900$	summer, spring	showers, long rain
Pluvial	< 1200	summer	long rain

5 Results

This section presents the results of the multi-scale evaluation framework (see Section 4) for several metrics related to the occurrence process of the precipitation events, daily amounts, and precipitation extremes. Summary assessments are provided, with several statistics provided for all the spatial scales of interest.

The precipitation observations are split into two sets: (1) 45 years randomly chosen among the period 1930-2014 are used to estimate the parameters and (2) the 40 remaining years are used to evaluate the performance of the models. This separation between an estimation set and a validation set is crucial to test the ability of the model to adequately represent the statistical properties of events which have not been used during the fitting procedure. In this study, the multi-scale evaluation is only applied to the 40-year validation set.

5.1 Parameter estimation and generation of scenarios

The different model parameters are estimated with the 45-year estimation set of observations, following the methodology described in section 3.3, except for the ξ parameter of the E-GPD which is estimated using all available precipitation data in Switzerland. This approach ensures that robust estimates are obtained for this parameter, which is crucial in our context since extreme simulated precipitation amounts are highly sensitive to the ξ parameter.

For GWEX, the estimation of the ξ parameter is performed at a daily scale. In order to highlight spatial patterns of ξ over Switzerland, we show the maps of the interpolated parameter estimates in Figure 4. Fat tails are obtained in the southern and eastern parts of the Aare River basin, particularly during spring and summer seasons. In the south of Switzerland, a region with high estimates ($\xi \sim 0.2$), highlighted in red, is obtained for the summer and autumn seasons. These high ξ estimates are consistent with the presence of strong convective storms in this mountainous region during this period of the year (Rudolph and Friedrich, 2012).

For GWEX_Disag, the regionalization method is applied at a 3-day scale (see Figure 5). The resulting estimates are similar to the ones obtained at a daily scale. However, note that the very high estimates obtained during the summer season at a daily scale are lower at a 3-day scale. This seems to confirm the interpretation of these high ξ estimates, i.e. the relationship between summer convective storms and high ξ estimates is not as strong at a 3-day scale, since storms of this type usually have a shorter duration. Note that non-zero ξ estimates in Figures 4 and 5 (in green, yellow and red) indicate that low-tailed distributions lead to an underestimation of extreme precipitation in these regions.

Figure 6 compares empirical and fitted distributions (mixture of exponentials and E-GPD) at a daily scale, for three illustrative stations and for the months of January, April, July and October. Both distributions fit the observed precipitation amounts reasonably well. Concerning the highest precipitation intensities, it is hard to draw conclusions on a significant over/underestimation. Indeed, local assessments of precipitation extremes are often inconclusive due to insufficient information on the distribution tails (Papalexiou and Koutsoyiannis, 2013).

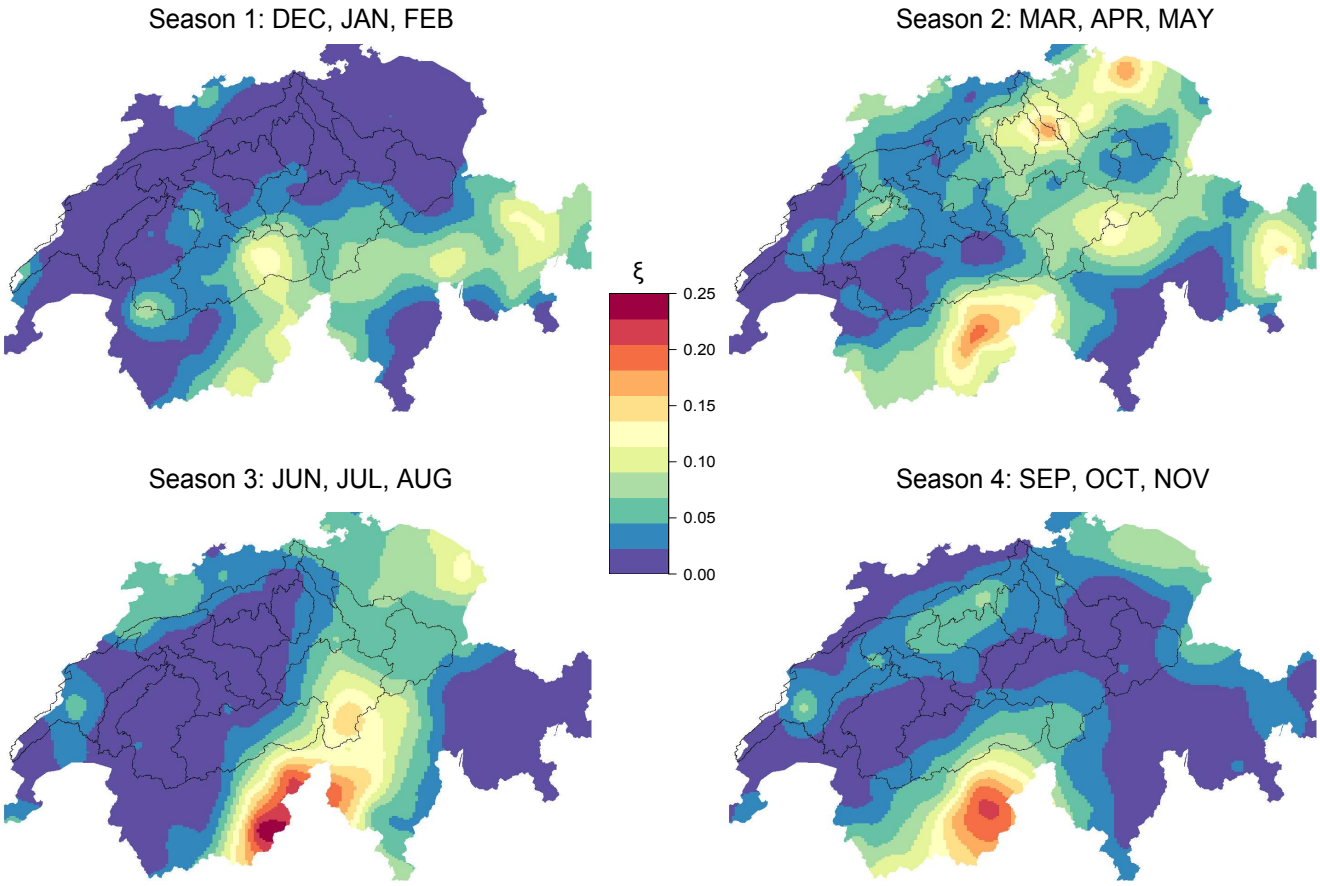


Figure 4. Regionalized ξ parameters at a daily scale, for the different seasons. Here, we present the spatial interpolation of at-site estimates for a better readability of their variability.

For each multi-site precipitation model investigated in this paper (Wilks, Wilks_EGPD, GWEX and GWEX_Disag), we generate 100 daily precipitation scenarios with these parameter estimates, each scenario having a length of 100 years. These scenarios are compared to the precipitation observed for the 40-year validation period.

5.2 Occurrence process

The monthly number of wet days obtained from observed and simulated precipitation data are compared in Figure 7. The average number of wet days is adequately reproduced by all models, with approximately 30% of cases with 'poor' performance. These 'poor' performance cases seem to occur mainly during the winter and spring seasons. The standard deviation of the monthly number of wet days indicates the inter-annual variability of this metric. While the magnitudes of the standard deviations from the simulated precipitation roughly match the corresponding observed standard deviations, it seems that the highest observed variabilities are underestimated by all the models, most markedly by the Wilks model.

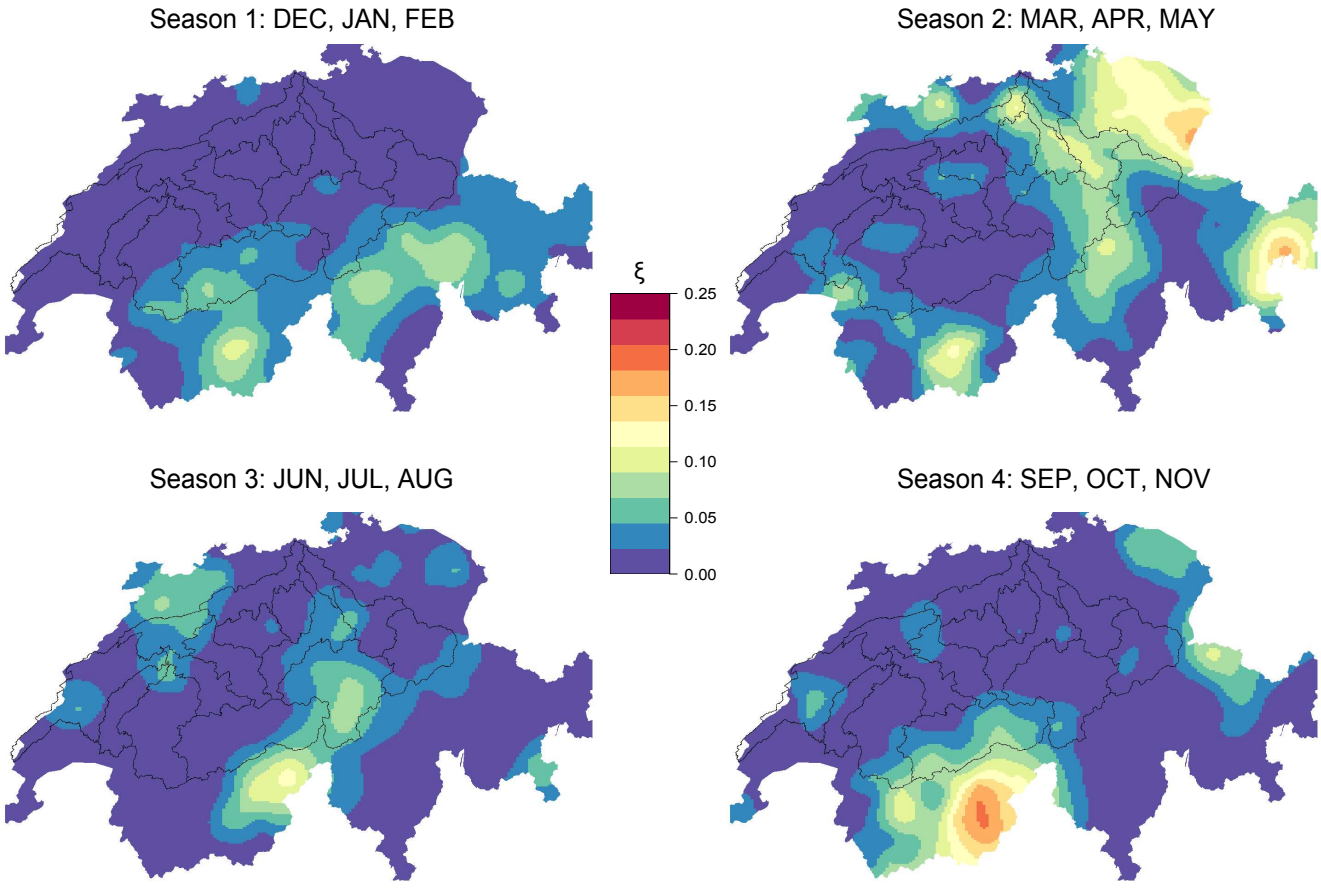


Figure 5. Regionalized ξ parameters at a 3-day scale, for the different seasons. Here, we present the spatial interpolation of at-site estimates for a better readability of their variability.

Figures 8 and 9 show the distributions of observed and simulated dry and wet spells, respectively, for the three illustrative stations. Concerning the distributions of dry spell lengths, the Wilks_EGPD, GWEX and GWEX_Disag models lead to adequate performance, the performance being classified as 'good' in 48%, 48% and 49% of the cases, respectively. The performance of the Wilks model is slightly lower because of an imprecise reproduction of the frequency of the shortest dry spells. This difference in performance is explained by the order of the Markov chain used to simulate the transitions between dry and wet states, which is the only difference between the occurrence processes of Wilks and Wilks_EGPD or GWEX. The 4-order Markov chain of the Wilks_EGPD and GWEX models seems to provide a more adequate representation of these transitions than the first-order Markov chain of the Wilks model, confirming previous findings (Srikanthan and Pegram, 2009).

The frequencies of wet spell lengths are adequately reproduced by the Wilks, Wilks_EGPD and GWEX models, with more than 50% of 'good' performance. The lower overall performance of GWEX_Disag for this metric is due to a slight underestimation of the longest wet spells for some stations (which is however not the case for the stations shown in Fig. 9).

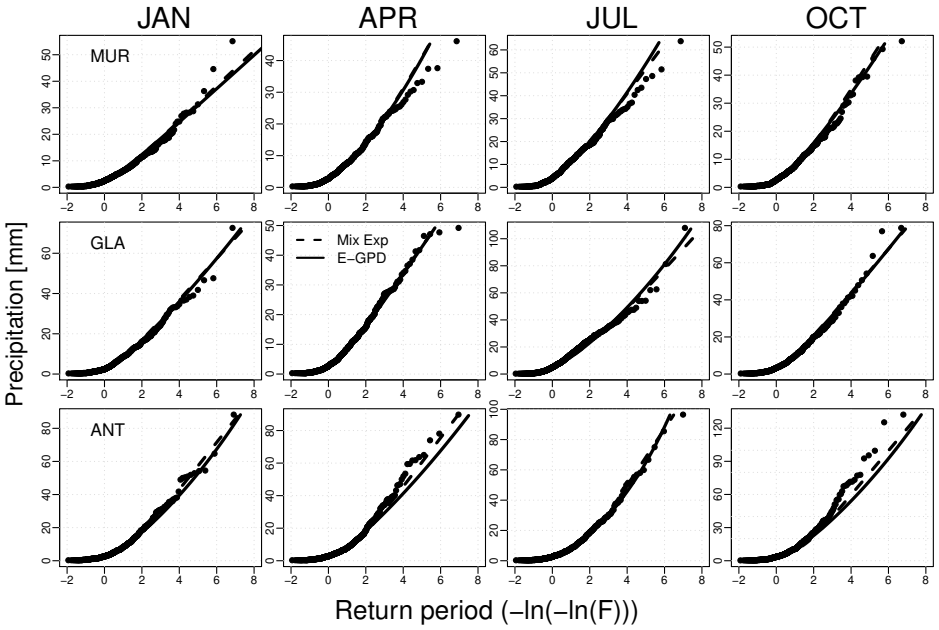


Figure 6. Empirical and fitted distributions (dashed curves for mixture of exponentials and solid curves for E-GPD) at a daily scale, for the three illustrative stations and for the months of January, April, July and October.

5.3 Inter-site correlations of precipitation amounts

Figure 10 compares observed and simulated inter-site correlations for the different model versions. Unlagged cross-correlations, which represent the spatial dependence, are close to the 1:1 diagonal line, as expected given that these correlations are explicitly taken into account by all model versions. However, a slight underestimation can be observed, especially concerning correlations above 0.8. This underestimation is a side-effect of the transformation applied to obtain a positive-definite matrix (see section 3.3).

An adequate reproduction of lag-1 inter-site correlations is important for the reproduction of persistent precipitation events. Simulated lag-1 cross-correlations are close to 0 for the Wilks and Wilks_EGPD models, as expected given that these versions ignore the temporal dependence. Consequently, these two model versions significantly underestimate observed lag-1 cross-correlations, which range between 0 and 0.4. Concerning GWEX, lag-1 serial autocorrelations at the stations (black points in the bottom plots) are perfectly aligned along the 1:1 line, as expected given that they are explicitly fitted by the MAR(1) process. Simulated and observed lag-1 cross-correlations are roughly in agreement, though the largest observed cross-correlations are underestimated. This is also the case to a lesser extent for GWEX_Disag. However, the agreement between observed and simulated cross-correlations is much stronger.

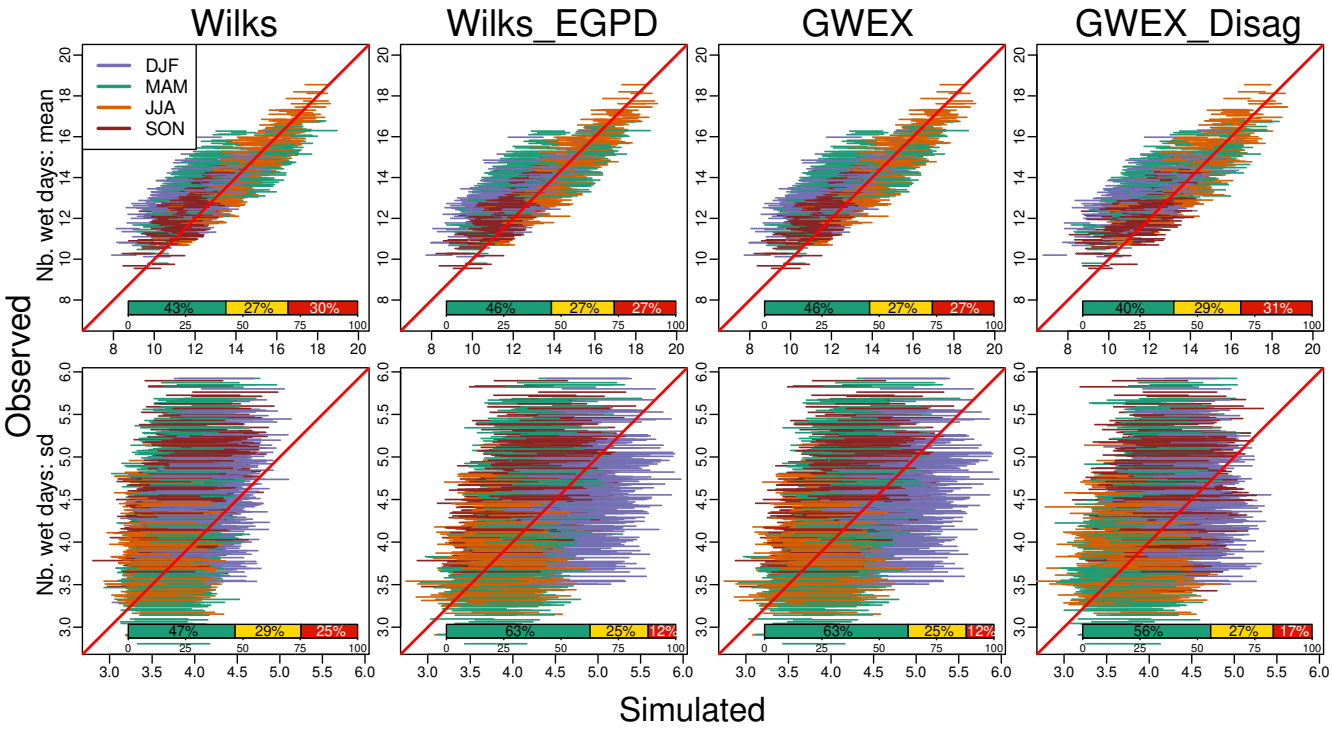


Figure 7. At-site number of wet days for all sites and months: inter-annual mean and standard deviation (sd). The 90% probability limits are shown for the different seasons. Overall performance is represented by the indicated percentages of 'good', 'fair' and 'poor' performance for all sites and months ($105 \times 12 = 1260$ cases).

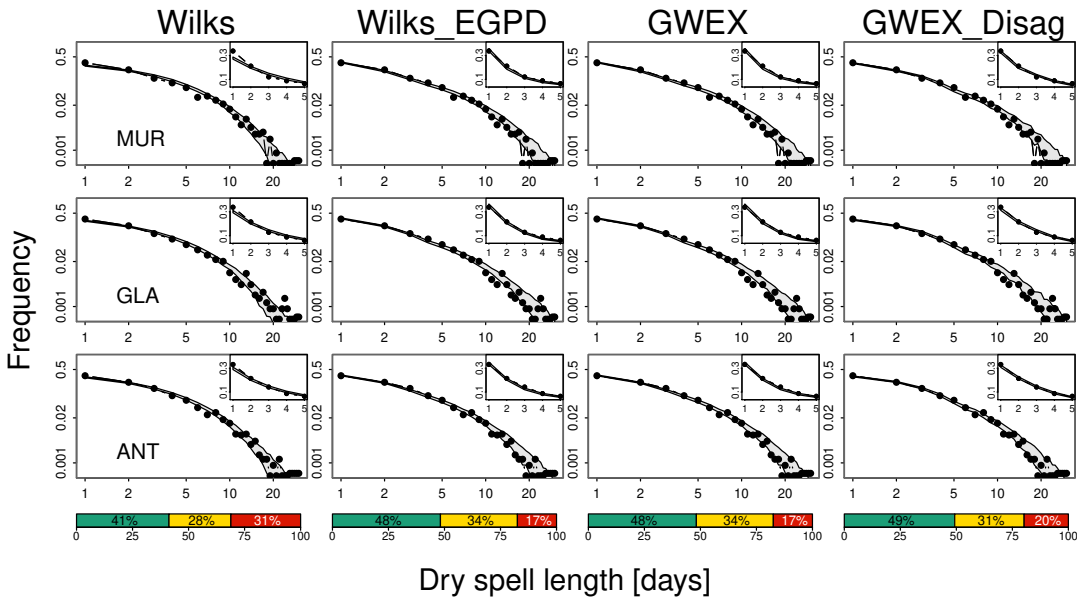


Figure 8. Distribution of dry spell lengths at the stations: The 90% probability limits are shown. Overall performance is represented by the indicated percentages of 'good', 'fair' and 'poor' performance for all sites. Inset plots provide a zoom for durations of 1 to 5 days.

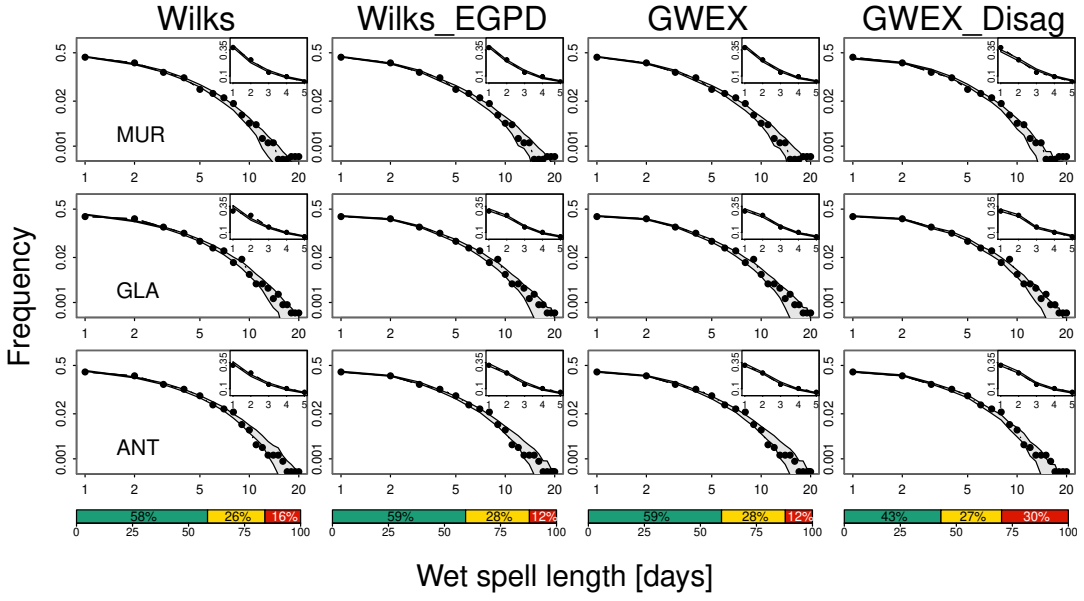


Figure 9. Distribution of wet spell lengths at the stations: The 90% probability limits are shown. Overall performance is represented by the indicated percentages of 'good', 'fair' and 'poor' performance for all sites. Inset plots provide a zoom for durations of 1 to 5 days.

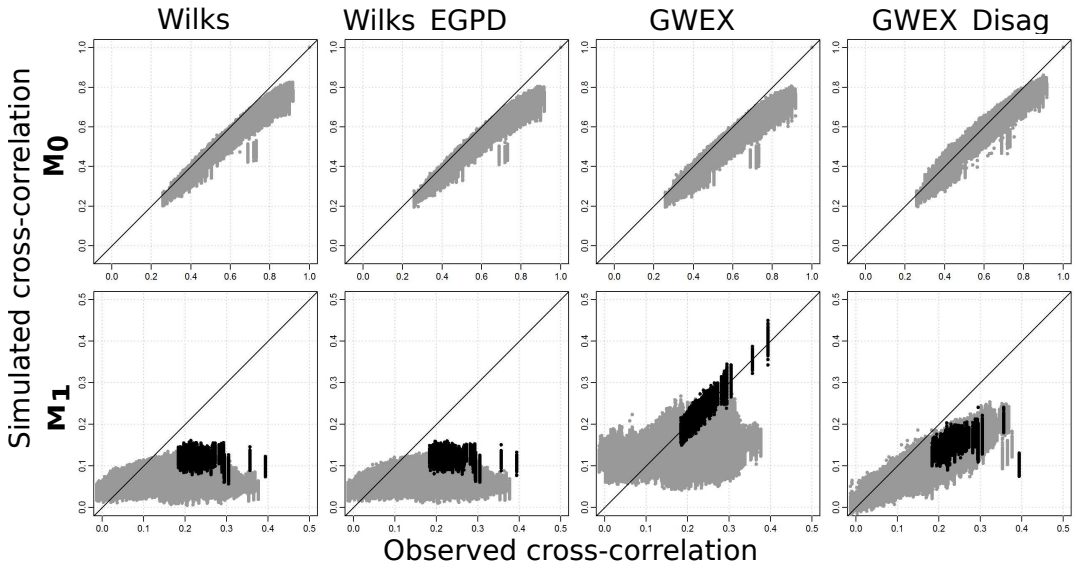


Figure 10. Comparison of unlagged inter-site correlations ($\mathbf{M}_0$) and lag-1 inter-site correlations ($\mathbf{M}_1$) in observed and simulated precipitation series, for the winter (DJF) and summer (JJA) seasons and for the different model versions considered. Black points indicate lag-1 serial autocorrelations at the stations.

5.4 Daily amounts

The reproduction of precipitation amounts at a daily scale is assessed in Figure 11, for all spatial scales and months. For all models, we obtain a reasonable agreement between observed and simulated average daily amounts (90% limits close to the 1:1 line), with more than 40% of 'good' cases and less than 30% of 'poor' cases. The standard deviations of these daily amounts are also adequately reproduced (Fig. 11, bottom plots).

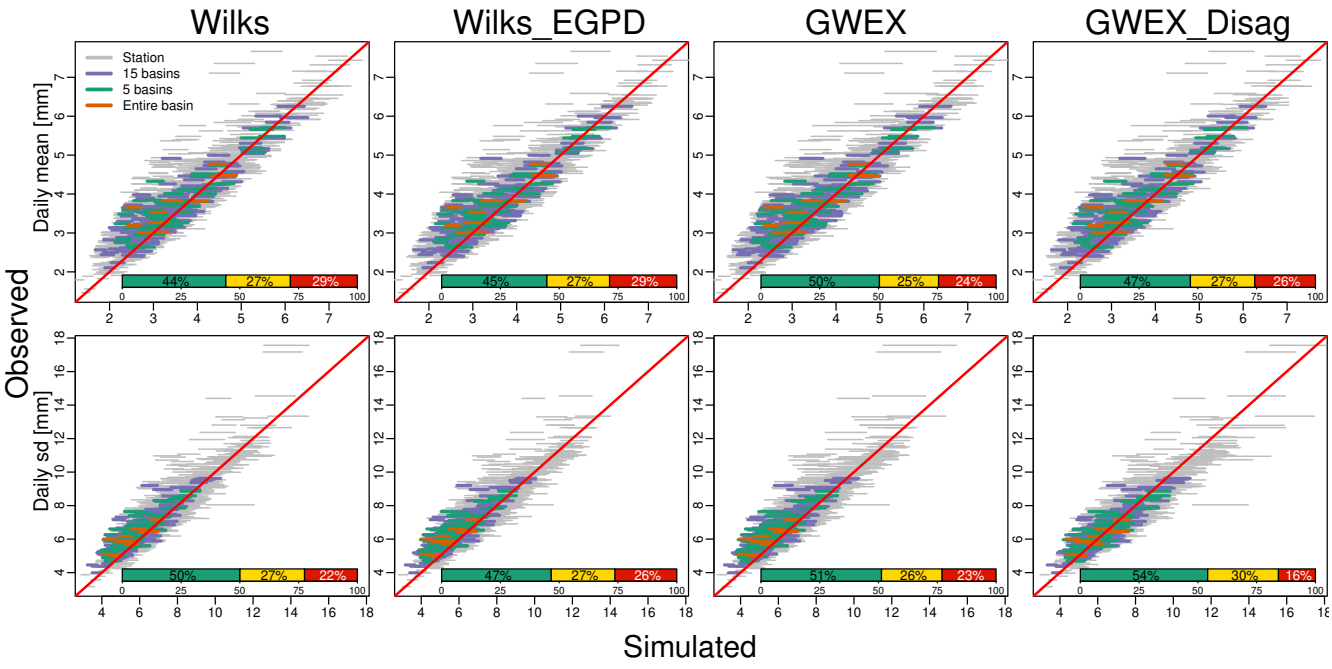


Figure 11. Daily amounts for all spatial scales and months: inter-annual mean (top) and standard deviation (sd, bottom). The 90% probability limits are shown. Overall performance is represented by the indicated percentages of 'good', 'fair' and 'poor' performance for all spatial scales and months.

5.5 Extreme precipitation amounts

Figures 12 and 13 show the relative differences, expressed as a percentage, between observed and simulated 10-year and 50-year return periods, at daily and 3-day scales, respectively, for all spatial scales. The percentiles corresponding to these return periods are estimated empirically using the Gringorten formula (Gringorten, 1963). These figures provide an overview of model performance regarding extreme precipitation amounts.

At the daily scale (Figure 12), there is no major difference in performance between the four models. For the 10 years and 50-year return periods, the number of 'poor' performance cases is below 20% for all models. The relative differences are globally centered around zero, which means that the mixture of exponentials (Wilks model) and the E-GPD (Wilks_EGPD, GWEX and GWEX_Disag models) all produce reasonable performance at this temporal scale. However, if we compare the 50-year return periods simulated by the Wilks and Wilks_EGPD models, we note an increase of 10% of 'good' performance cases (from 65% to 75%), which can be explained by a slight underestimation of the largest maxima with Wilks, for some stations.

Comparing Wilks_EGPD and GWEX, the scores are almost identical, which suggests that the tail dependence introduced by the Student copula in GWEX does not produce a significant improvement for the reproduction of extremes. However, if we focus on the largest spatial scales (at the basins), and in particular on the entire Aare River basin (orange lines), it seems that the slight underestimation of the 50-year return periods obtained with Wilks_EGPD is reduced thanks to this tail dependence. GWEX_Disag also reproduces adequately the largest precipitation amounts at all spatial scales, even if a slight overestimation of the maxima at the largest spatial scales can be suspected. Nevertheless, this performance shows that the disaggregation process leads to an adequate reproduction of the daily maxima.

At the 3-day scale (Figure 13), the underestimation of the maxima by Wilks and Wilks_EGPD is clear at all spatial scales. GWEX does not suffer from the same shortcomings, which means that the MAR(1) process (Eq. 10) improves the temporal structure of the largest 3-day precipitation amounts. GWEX_Disag being fitted at a 3-day scale, this model logically leads to an adequate reproduction of extreme 3-day precipitation amounts. The strategy consisting in simulating 3-day precipitation amounts, which are then disaggregated at a daily scale, presents several advantages:

- The model being fitted at a 3-day scale, 3-day maxima are adequately reproduced.
- As the method of fragments uses observed 3-day temporal structures to disaggregate 3-day amounts, the daily amounts resulting from a generated 3-day maxima are physically plausible. In particular, the temporal and spatial structures of large and persistent observed precipitation events are used, which ensures consistency between the generated extreme events at the daily and 3-day scales.

GWEX and GWEX_Disag both adequately reproduce extreme precipitation amounts at daily and 3-day scales, as well as at all spatial scales. As indicated above, these models will be used to generate long precipitation scenarios, which will feed a hydrological model in order to produce flood scenarios. Ultimately, the reproduction of the flood properties using GWEX and GWEX_Disag will indicate which model is the most adequate. Since they correspond to the same model version fitted at daily and 3-day scale, respectively, we can expect that resulting floods will have slightly different properties.

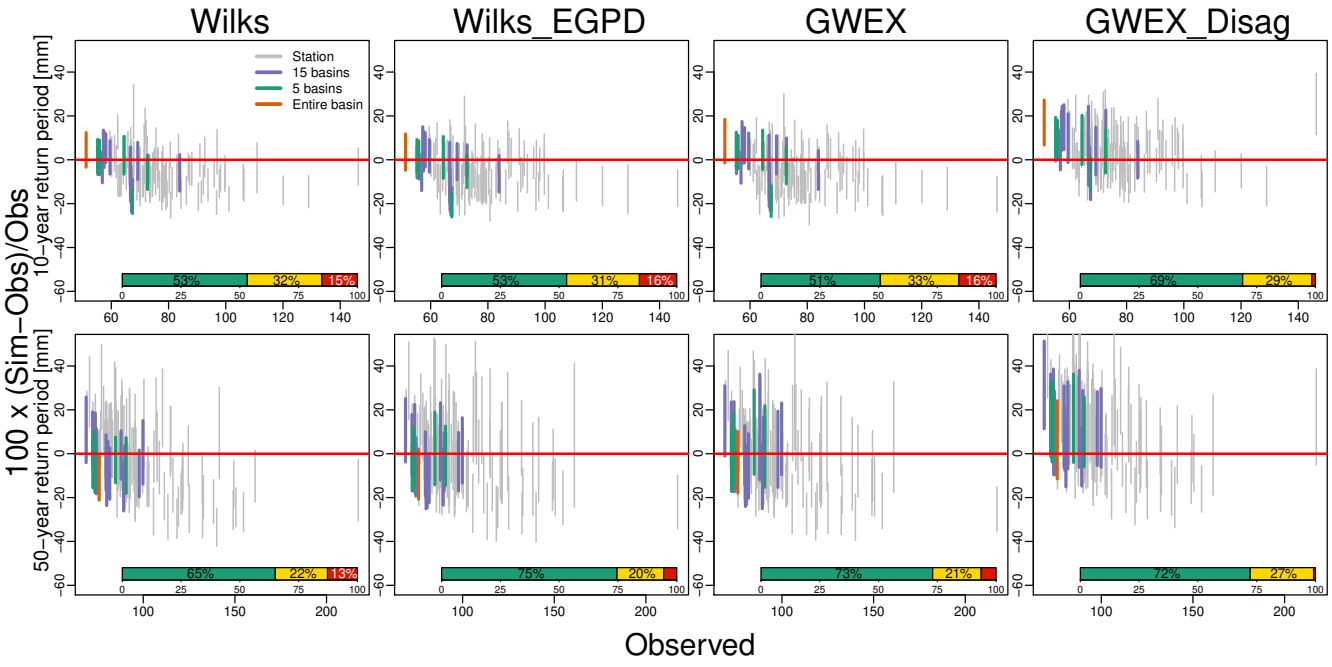


Figure 12. Daily annual maxima for all spatial scales: Relative differences, expressed as a percentage, between observed and simulated 10-year (top plots) and 50-year (bottom plots) return periods. The 90% probability limits are shown. Overall performance is represented by the indicated percentages of 'good', 'fair' and 'poor' performance for all spatial scales.

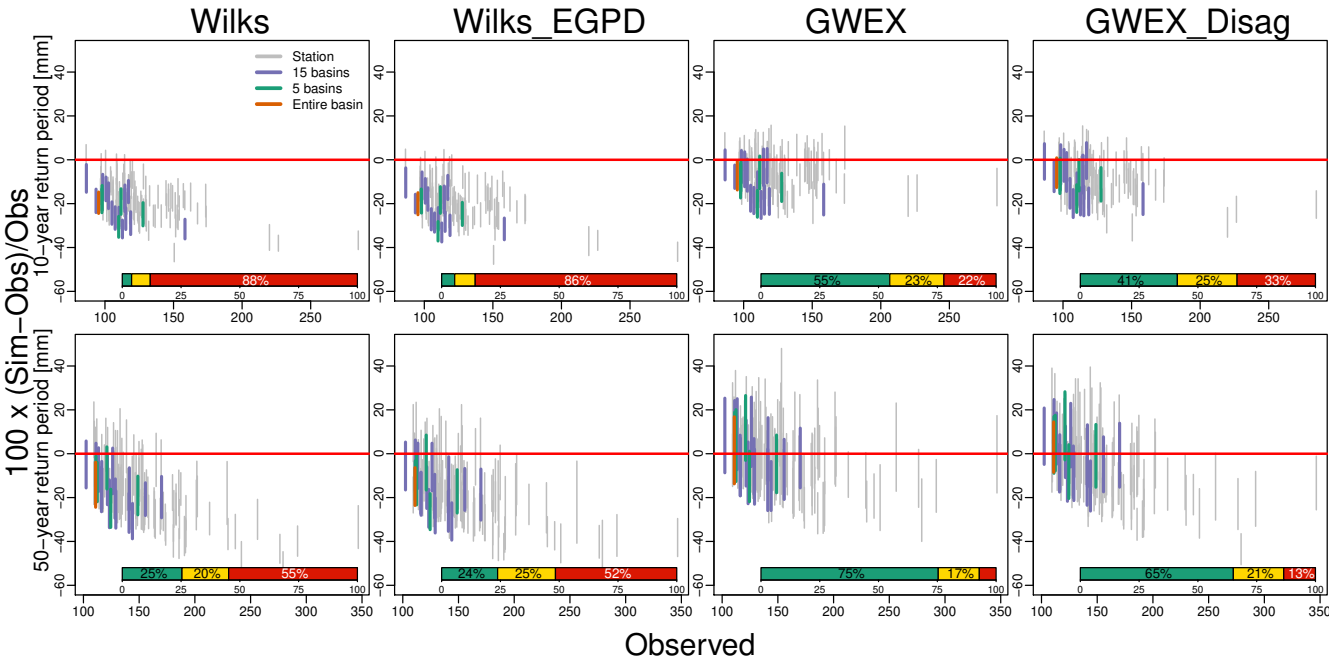


Figure 13. 3-day annual maxima for all spatial scales: Relative differences, expressed as a percentage, between observed and simulated 10-year (top plots) and 50-year (bottom plots) return periods. The 90% probability limits are shown. Overall performance is represented by the indicated percentages of 'good', 'fair' and 'poor' performance for all spatial scales.

6 Conclusions and outlook

Precipitation models are usually developed for the purpose of risk assessment in relation to natural hazards (e.g. droughts, floods). Most existing precipitation models aim at reproducing a wide range of statistical properties of precipitation, at different scales, in order to be used as a general tool in different contexts. In this study, our main objective was to provide a precipitation generator that could be used together with a hydrological model for the evaluation of extreme flooding events in a region covering approximately half of Switzerland. As a consequence, we were especially interested in the reproduction of extreme precipitation amounts at medium to large spatial scales. As the daily and 3-day precipitation amounts are a major determinant of flood magnitudes in large Swiss basins (Froidevaux et al., 2015), an adequate reproduction of precipitation at these time scales was also required.

In this paper, we considered different multi-site precipitation models targeting the reproduction of extreme amounts at multiple temporal (daily, 3-day) and spatial scales. Different extended versions of the model introduced by Wilks (Wilks, 1998) have been proposed. A first direct extension, Wilks_EGPD, considers a Markov chain of order 4 instead of order 1) for the at-site occurrence process. Furthermore, taking advantage of recent advances regarding extreme precipitation, a heavy-tailed distribution (instead of a mixture of exponential distributions), the E-GPD, is applied to the precipitation intensities at each station. Two important extensions of Wilks_EGPD, named GWEX and GWEX_Disag, are then considered. In GWEX model, temporal and spatial dependencies of the occurrence and intensity process are introduced using the copula theory and a multivariate autoregressive process. A second version, GWEX_Disag, applies the same model, but at a 3-day scale. The 3-day simulated amounts are then disaggregated using an adaptation of the method of fragments (Wójcik and Buishand, 2003).

In this study, we support the use of a systematic evaluation framework. The CASE framework proposed by Bennett et al. (2017) provides a useful tool in this respect, making it possible to fairly compare performance between precipitation models. Regarding the reproduction of extreme precipitation, evaluations until now have usually been qualitative (e.g. interpretations based on one or two examples) and limited in terms of spatial scales (often only at the stations). The evaluation of extreme precipitation amounts proposed in this paper is multi-scale in time (daily and 3-day scale) and space (at the stations, for two different divisions of the study area into sub-basins, and for the entire Aare River basin).

The different multi-site precipitation models have been applied to 105 stations located in Switzerland. A multi-scale evaluation led to the following conclusions:

- A fourth-order Markov chain outperforms a first-order Markov chain for the transitions between dry and wet states, notably for the reproduction of dry spell lengths.
- At the scale of the stations, daily amounts (average, standard deviations and extremes) are reasonably well reproduced by all the models.
- With only three parameters, the E-GPD provides a parsimonious and flexible representation of the whole of precipitation amounts. Its GPD tail is in agreement with recent results showing that extreme precipitation amounts must be modeled by heavy-tailed distributions (Papalexiou and Koutsoyiannis, 2013; Serinaldi and Kilsby, 2014a). Furthermore, robust

estimates of the parameter controlling the heaviness of the distribution tail are obtained using a regionalization method. In our study area, the E-GPD does not bring a significant improvement of the performance compared to the mixture of exponential distributions. However, the general framework proposed in this paper can be applied to very distinct precipitation regimes and the possible heavy tail of the E-GPD might be valuable in other areas.

- At a 3-day scale, precipitation extremes are severely underestimated by Wilks and Wilks_EGPD. This underestimation can be explained by an incorrect representation of the persistence by these models.

- GWEX and GWEX_Disag adequately reproduce extreme precipitation amounts at daily and 3-day scales, and at all spatial scales. These models are deemed adequate for the evaluation of extreme flood events.

Future research will investigate if the floods simulated by a hydrological model using the generated precipitation scenarios have statistical properties in agreement with observed floods. An extensive investigation is currently underway with a distributed version of the HBV hydrological model, applied to 87 sub-basins of the whole study area and using precipitation scenarios produced by GWEX as inputs. This hydrological evaluation of our weather scenarios will be presented in future publications.

Appendix A: Temporal disaggregation from a 3-day scale to a daily scale

For a 3-day period $\mathbf{D} = \{d, d+1, d+2\}$ starting on a day d, the observed and simulated precipitation amounts at a station k are denoted by $Y_{\mathbf{D}}(k)$ and $\tilde{Y}_{\mathbf{D}}(k)$, respectively. We want to disaggregate the simulated 3-day amount for the period $\tilde{\mathbf{D}} = \{\tilde{d}, \tilde{d}+1, \tilde{d}+2\}$. This disaggregation is achieved in the following steps:

1. A set of observed 3-day sequences are retained as candidate periods $\mathbf{D}$ according to two criteria:

   - **Season:** Periods $\tilde{\mathbf{D}}$ and $\mathbf{D}$ must belong to the same season, as defined in Section 3.3.
   - **Mean intensity:** Simulated and observed precipitation fields must have the same order of magnitude. Let $q_{0.5}$, $q_{0.75}$, $q_{0.9}$ and $q_{0.99}$ denote the quantiles of the mean observed precipitation intensities over all the stations associated with probabilities 0.5, 0.75, 0.9, and 0.99, respectively. Observed and simulated 3-day periods are classified in 5 groups according to their mean intensity $\bar{\mathbf{Y}} = \frac{1}{n}\sum_k Y_{\mathbf{D}}(k)$: dry periods ($\bar{\mathbf{Y}} < q_{0.5}$), moderately wet periods ($q_{0.5} \leq \bar{\mathbf{Y}} < q_{0.75}$), wet periods ($q_{0.75} \leq \bar{\mathbf{Y}} < q_{0.9}$), very wet periods ($q_{0.9} \leq \bar{\mathbf{Y}} < q_{0.99}$) and extremely wet periods ($q_{0.99} \geq \bar{\mathbf{Y}}$).

   This first selection of candidate periods aims at increasing the chance of retaining periods corresponding to similar meteorological events.

2. For each observed 3-day candidate period $\mathbf{D}$, we compute the following score:

$$SCORE(\tilde{\mathbf{D}}, \mathbf{D}) = \sum_k \left| \frac{\tilde{Y}_{\tilde{d}-1}(k)}{\sum_k \tilde{Y}_{\tilde{d}-1}(k)} - \frac{Y_{d-1}(k)}{\sum_k Y_{d-1}(k)} \right| + \left| \frac{\tilde{Y}_{\mathbf{D}}(k)}{\sum_k \tilde{Y}_{\mathbf{D}}(k)} - \frac{Y_{\mathbf{D}}(k)}{\sum_k Y_{\mathbf{D}}(k)} \right|.$$

   This score measures the similarity between the simulated spatial field for the period $\tilde{Y}_{\mathbf{D}}(k)$ and the observed spatial field for the period $\tilde{\mathbf{D}}$ and also takes into account the similarity between the spatial fields for the previous days $\tilde{d}-1$ and $d-1$. Absolute differences between relative precipitation intensities are computed (the lowest scores are therefore obtained for spatial fields with similar shapes), among the observed periods corresponding to the same season and order of magnitude selected in the previous step.

3. For each simulated period $\tilde{\mathbf{D}}$, the observed precipitation fields corresponding to the 10 lowest scores are retained. For each station k, if a positive precipitation amount has been simulated ($\tilde{Y}_{\tilde{\mathbf{D}}}(k) > 0$), we look at the corresponding observed amount $Y_{\mathbf{D}}(k)$. If $Y_{\mathbf{D}}(k) = 0$, this observed period cannot be used to disaggregate $\tilde{Y}_{\tilde{\mathbf{D}}}(k)$ and we look at the next best observed field among the 10 selected fields. If the observed field contains a positive precipitation amount at this station ($Y_{\mathbf{D}}(k) > 0$), then we obtain the simulated daily amount for day $\tilde{d}$ as follows:

$$\tilde{Y}_{\tilde{d}}(k) = Y_d(k) \times \frac{\tilde{Y}_{\tilde{\mathbf{D}}}(k)}{Y_{\mathbf{D}}(k)}, \tag{A1}$$

with similar expressions for days $\tilde{d}+1$ and $\tilde{d}+2$. Simulated daily amounts correspond to the observed daily amounts, rescaled by the ratio between the simulated and observed 3-day amounts. The 3-day simulated amounts and observed temporal structures are thus preserved.

4. While the 3-day spatio-temporal consistency is generally conserved by applying the preceding steps, it can happen that the simulated 3-day amount is positive even though there is no positive precipitation among the 10 best 3-day observed fields. In this case, we seek similar observed amounts at this station only and randomly choose one 3-day period among the 10 best 3-day periods.

Acknowledgements. We gratefully acknowledge financial support for this study provided by the Swiss Federal Office for Environment (FOEN), the Swiss Federal Nuclear Safety Inspectorate (ENSI), the Federal Office for Civil Protection (FOCP) and the Federal Office of Meteorology and Climatology, MeteoSwiss, through the project EXAR ("Evaluation of extreme Flooding Events within the Aare-Rhine hydrological system in Switzerland").

References

Ailliot, P., Allard, D., Monbet, V., and Naveau, P.: Stochastic weather generators: an overview of weather type models, Journal de la Société Française de Statistique, 156, 101–113, 2015.

Allard, D. and Bourotte, M.: Disaggregating daily precipitations into hourly values with a transformed censored latent Gaussian process, Stochastic Environmental Research and Risk Assessment, 29, 453–462, 2015.

Baxevani, A. and Lennartsson, J.: A spatiotemporal precipitation generator based on a censored latent Gaussian field, Water Resources Research, 51, 4338–4358, 2015.

Bennett, B., Thyer, M., Leonard, M., Lambert, M., and Bates, B.: A comprehensive and systematic evaluation framework for a parsimonious daily rainfall field model, Journal of Hydrology, https://doi.org/10.1016/j.jhydrol.2016.12.043, In Press, 2017.

Bárdossy, A. and Pegram, G. G. S.: Copula based multisite model for daily precipitation simulation, Hydrology and Earth System Sciences, 13, 2299–2314, 2009.

Buishand, T. A.: Extreme rainfall estimation by combining data from several sites, Hydrological Sciences Journal, 36, 345–365, 1991.

Burn, D. H.: Evaluation of regional flood frequency analysis with a region of influence approach, Water Resources Research, 26, 2257–2265, 1990.

Burton, A., Kilsby, C. G., Fowler, H. J., Cowpertwait, P. S. P., and O'Connell, P. E.: RainSim: A spatial–temporal stochastic rainfall modelling system, Environmental Modelling & Software, 23, 1356–1369, 2008.

Burton, A., Fowler, H. J., Kilsby, C. G., and O'Connell, P. E.: A stochastic model for the spatial-temporal simulation of nonhomogeneous rainfall occurrence and amounts, Water Resources Research, 46, W11 501, 2010.

Cavanaugh, N. R., Gershunov, A., Panorska, A. K., and Kozubowski, T. J.: The probability distribution of intense daily precipitation, Geophysical Research Letters, 42, 2015GL063 238, 2015.

Chandler, R. E. and Wheater, H. S.: Analysis of rainfall variability using generalized linear models: A case study from the west of Ireland, Water Resources Research, 38, 1192, 2002.

Chen, J., Brissette, F. P., and Zhang, J. X.: A Multi-Site Stochastic Weather Generator for Daily Precipitation and Temperature, Transactions of the ASABE, pp. 1375–1391, 2014.

Dempster, A. P., Laird, N. M., and Rubin, D. B.: Maximum Likelihood from Incomplete Data via the EM Algorithm, Journal of the Royal Statistical Society. Series B (Methodological), 39, 1–38, 1977.

Evin, G., Blanchet, J., Paquet, E., Garavaglia, F., and Penot, D.: A regional model for extreme rainfall based on weather patterns subsampling, Journal of Hydrology, 541, Part B, 1185–1198, 2016.

Froidevaux, P.: Meteorological characterisation of floods in Switzerland, Ph.D. thesis, Geographisches Institut, University of Bern, 2014.

Froidevaux, P., Schwanbeck, J., Weingartner, R., Chevalier, C., and Martius, O.: Flood triggering in Switzerland: the role of daily to monthly preceding precipitation, Hydrology and Earth System Sciences, 19, 3903–3924, 2015.

Gringorten, I. I.: A plotting rule for extreme probability paper, Journal of Geophysical Research, 68, 813–814, 1963.

Hundecha, Y., Pahlow, M., and Schumann, A.: Modeling of daily precipitation at multiple locations using a mixture of distributions to characterize the extremes, Water Resources Research, 45, W12 412, 2009.

Keller, D. E., Fischer, A. M., Frei, C., Liniger, M. A., Appenzeller, C., and Knutti, R.: Implementation and validation of a Wilks-type multisite daily precipitation generator over a typical Alpine river catchment, Hydrology and Earth System Sciences, 19, 2163–2177, 2015.

Kleiber, W., Katz, R. W., and Rajagopalan, B.: Daily spatiotemporal precipitation simulation using latent and transformed Gaussian processes, Water Resources Research, 48, W01 523, 2012.

Klemeš, V.: Tall Tales about Tails of Hydrological Distributions. I., Journal of Hydrologic Engineering, 5, 227–231, 2000a.

Klemeš, V.: Tall Tales about Tails of Hydrological Distributions. II., Journal of Hydrologic Engineering, 5, 232–239, 2000b.

Lafaysse, M., Hingray, B., Mezghani, A., Gailhard, J., and Terray, L.: Internal variability and model uncertainty components in future hydrometeorological projections: The Alpine Durance basin, Water Resources Research, 50, 3317–3341, 2014.

Lamb, R., Faulkner, D., Wass, P., and Cameron, D.: Have applications of continuous rainfall–runoff simulation realized the vision for process-based flood frequency analysis?, Hydrological Processes, 30, 2463–2481, 2016.

Leblois, E. and Creutin, J.-D.: Space-time simulation of intermittent rainfall with prescribed advection field: Adaptation of the turning band method, Water Resources Research, 49, 3375–3387, 2013.

Leonard, M., Lambert, M. F., Metcalfe, A. V., and Cowpertwait, P. S. P.: A space-time Neyman–Scott rainfall model with defined storm extent, Water Resources Research, 44, W09 402, 2008.

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., Rust, H. W., Sauter, T., Themeßl, M., Venema, V. K. C., Chun, K. P., Goodess, C. M., Jones, R. G., Onof, C., Vrac, M., and Thiele-Eich, I.: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, Reviews of Geophysics, 48, RG3003, 2010.

McNeil, A. J., Frey, R., and Embrechts, P.: Quantitative Risk Management - Concepts, Techniques, and Tools, Princeton University Press, Princeton, N.J, 2005.

McRobie, F. H., Wang, L.-P., Onof, C., and Kenney, S.: A spatial-temporal rainfall generator for urban drainage design, Water Science and Technology: A Journal of the International Association on Water Pollution Research, 68, 240–249, 2013.

Mehrotra, R. and Sharma, A.: Preserving low-frequency variability in generated daily rainfall sequences, Journal of Hydrology, 345, 102–120, 2007a.

Mehrotra, R. and Sharma, A.: A semi-parametric model for stochastic generation of multi-site daily rainfall exhibiting low-frequency variability, Journal of Hydrology, 335, 180–193, 2007b.

Mehrotra, R. and Sharma, A.: Development and Application of a Multisite Rainfall Stochastic Downscaling Framework for Climate Change Impact Assessment, Water Resources Research, 46, W07 526, 2010.

Mehrotra, R., Srikanthan, R., and Sharma, A.: A comparison of three stochastic multi-site precipitation occurrence generators, Journal of Hydrology, 331, 280–292, 2006.

Mezghani, A. and Hingray, B.: A combined downscaling-disaggregation weather generator for stochastic generation of multisite hourly weather variables over complex terrain: Development and multi-scale validation for the Upper Rhone River basin, Journal of Hydrology, 377, 245–260, 2009.

Naveau, P., Huser, R., Ribereau, P., and Hannart, A.: Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection, Water Resources Research, 52, 2753–2769, 2016.

Papalexiou, S. M. and Koutsoyiannis, D.: Battle of extreme value distributions: A global survey on extreme daily rainfall, Water Resources Research, 49, 187–201, 2013.

Papalexiou, S. M., Koutsoyiannis, D., and Makropoulos, C.: How extreme is extreme? An assessment of daily rainfall distribution tails, Hydrology and Earth System Sciences, 17, 851–862, 2013.

Papastathopoulos, I. and Tawn, J. A.: Extended generalised Pareto models for tail estimation, Journal of Statistical Planning and Inference, 143, 131–143, 2013.

Rasmussen, P. F.: Multisite precipitation generation using a latent autoregressive model, Water Resources Research, 49, 1845–1857, 2013.

Rayner, D., Achberger, C., and Chen, D.: A multi-state weather generator for daily precipitation for the Torne River basin, northern Sweden/western Finland, Advances in Climate Change Research, 7, 70–81, 2016.

Rebonato, R. and Jaeckel, P.: The Most General Methodology to Create a Valid Correlation Matrix for Risk Management and Option Pricing Purposes, SSRN Scholarly Paper ID 1969689, Social Science Research Network, Rochester, NY, 2011.

Rousseeuw, P. J. and Molenberghs, G.: Transformation of non positive semidefinite correlation matrices, Communications in Statistics - Theory and Methods, 22, 965–984, 1993.

Rudolph, J. V. and Friedrich, K.: Seasonality of Vertical Structure in Radar-Observed Precipitation over Southern Switzerland, Journal of Hydrometeorology, 14, 318–330, 2012.

Serinaldi, F. and Kilsby, C. G.: Rainfall extremes: Toward reconciliation after the battle of distributions, Water Resources Research, 50, 336–352, 2014a.

Serinaldi, F. and Kilsby, C. G.: Simulating daily rainfall fields over large areas for collective risk estimation, Journal of Hydrology, 512, 285–302, 2014b.

Sharif, M. and Burn, D. H.: Improved K -Nearest Neighbor Weather Generating Model, Journal of Hydrologic Engineering, 12, 2007.

Srikanthan, R. and Pegram, G. G. S.: A nested multisite daily rainfall stochastic generation model, Journal of Hydrology, 371, 142–153, 2009.

Thompson, C. S., Thomson, P. J., and Zheng, X.: Fitting a multisite daily rainfall model to New Zealand data, Journal of Hydrology, 340, 25–39, 2007.

Vischel, T., Lebel, T., Massuel, S., and Cappelaere, B.: Conditional simulation schemes of rain fields and their application to rainfall–runoff modeling studies in the Sahel, Journal of Hydrology, 375, 273–286, 2009.

Vrac, M. and Naveau, P.: Stochastic downscaling of precipitation: From dry events to heavy rainfalls, Water Resources Research, 43, W07 402, 2007.

Wilks, D. S.: Multisite generalization of a daily stochastic precipitation generation model, Journal of Hydrology, 210, 178–191, 1998.

Wójcik, R. and Buishand, T.: Simulation of 6-hourly rainfall and temperature by two resampling schemes, Journal of Hydrology, pp. 69–80, 2003.

Yates, D., Gangopadhyay, S., Rajagopalan, B., and Strzepek, K.: A technique for generating regional climate scenarios using a nearest-neighbor algorithm, Water Resources Research, 39, 1199, 2003.

10.3.2 Stochastic generators of multi-site daily temperature

Stochastic generators of multi-site daily temperature: Comparison of performance in various applications

Guillaume Evin Anne-Catherine Favre
Benoit Hingray
Univ. Grenoble Alpes, CNRS, IRD, Grenoble INP*, IGE,
F-38000 Grenoble, France

January 31, 2018

Abstract

We present a multi-site stochastic model for the generation of average daily temperature, which includes a flexible parametric distribution and a multivariate autoregressive process. Different versions of this model are applied on a set of 26 stations located in Switzerland. The importance of specific statistical characteristics of the model (seasonality, marginal distributions of standardized temperature, spatial and temporal dependence) is discussed. In particular, the proposed marginal distribution is shown to improve the reproduction of extreme temperatures (minima and maxima). We also demonstrate that the frequency and duration of cold spells and heat waves are dramatically underestimated when the autocorrelation of temperature is not taken into account in the model. An adequate representation of these characteristics can be crucial depending of the field of application and we discuss potential implications in different contexts (agriculture, forestry, hydrology, human health).

*Institute of Engineering Univ. Grenoble Alpes

1 Introduction

Risk assessment studies often apply stochastic generators in order to generate long series of different meteorological variables, such as precipitation, temperature, solar radiation or wind (e.g., Richardson, 1981; Wilks and Wilby, 1999; Apipattanavis et al., 2007; Leander and Buishand, 2009; Steinschneider and Brown, 2013; Chen et al., 2014; Li, 2014; Breinl et al., 2015). These long scenarios are then used as inputs of environmental models. This type of applications can be found in various fields, for example in agriculture (Stöckle et al., 2003; Romero et al., 2009; Deryng et al., 2014; Yang et al., 2014), in biogeochemistry (Kucharik et al., 2000; Krinner et al., 2005; Guenet et al., 2013), for studies on human health (Boumans et al., 2014) or for the impact of climate change (Wilks, 1992; Deryng et al., 2014).

This paper focus on the stochastic generation of temperature, which can have environmental impacts at several levels, as illustrated by the following examples:

- Extreme temperatures can be directly responsible of natural disasters (wildfires), economical loss (crop loss) and human loss (heat wave).
- Evapo-transpiration is directly related to temperature and affects the availability of water resources.
- When temperature becomes negative, precipitation is accumulated as snow. The proportion of precipitation falling as snow instead of rainfall is crucial in some environmental applications, for example in hydrology. Indeed, in many catchments, a number of floods are directly related to the snow cover (spring floods, rain-on-snow).

A popular structure, originally proposed by Wilks (1998) to model precipitation at multiple sites, is applied to minimum and maximum daily temperatures by Wilks and Wilby (1999). Several models adopt this approach to generate simultaneously precipitation and temperature time series (Chen et al., 2014; Baigorria, 2014). Alternatively, Generalized Linear Model (GLM) can be applied to link temperature or other meteorological variables to climatic covariates (Chandler and Wheater, 2002; Mezghani and Hingray, 2009; Asong et al., 2016; Chandler, 2016). Resampling methods (e.g. K-nearest neighbors) is also an option which has been investigated by King et al. (2014, 2015). In

a recent paper, Erhardt et al. (2015) model the spatial structure of daily mean temperature using vine copulas.

In this study, we extend the structure proposed by Wilks (1999). In particular, we introduce a flexible parametric distribution, the Skew Exponential Power (SEP) distribution, which can reproduce various levels of asymmetry and flatness. A multivariate autoregressive process represents simultaneously temporal and spatial inter-site correlations. Using Wilks' structure, this multi-site stochastic model is able to represent both spatial and temporal dependences over large areas, possibly with high differences of elevation. It is designed to offer a flexible and robust framework which can adapt very different types of climate and reproduce various temperature features. The four main components of this model can be quickly described as:

- **Non-stationarity:** As temperature data exhibit a clear increase of the average temperature in Switzerland, a seasonal linear trend is introduced.
- **Seasonality:** The seasonal cycles of temperature (mean and standard deviations) are modeled by non-parametric functions.
- **Marginal distributions:** At each station, the distribution modeling standardized temperature data is the SEP distribution. This flexible 4-parameter distribution can describe different shapes and levels of asymmetry (skewness) and flatness (kurtosis).
- **Multivariate temporal and spatial dependence:** The dependence between the temperature data at different stations, as well as the correlation between temperature at days d and $(d + 1)$, is modelled by a Multivariate Autoregressive Process (MAR) process of order one.

This paper intends to demonstrate the importance of specific components of the temperature model for the reproduction of different statistical features (e.g. seasonal cycles, inter-station dependence, etc.). We discuss how these statistical properties are crucial in specific applications (agriculture, forestry, hydrology, human health). For example, while most of the past studies generally focus on spatial aspects of temperature (e.g. inter-site correlations), this paper also appraise the reproduction of extreme temperature events. The SEP distribution is shown to improve the reproductions of extreme temperatures (minima and maxima). We also demonstrate that the frequency and duration of cold spells and heat waves are dramatically underestimated when

the autocorrelation of temperature is not taken into account in the model.

The global methodology is first described in Section 2, with a presentation of the data and study area and the features of the multi-site temperature model. Section 3 presents an application of this daily temperature model to 26 stations located in the Aare river catchment, in Switzerland. Section 4 concludes.

2 Material and methods

2.1 Data and study area

Figure 1 shows a map of Switzerland, with the location of 26 temperature stations. For these stations, long daily time series of observations are available during the period 1930-2014, with less than 3 years of missing data. In this study, weather scenarios are generated at these stations using a multi-site stochastic model.

For the sake of illustration, 6 temperature stations are selected among these 26 stations. Table 1 provides some characteristics for these stations (elevation, winter temperature mean, winter temperature minimum, summer temperature mean and summer temperature maximum). Stations in the Swiss plateau (NEU at Neuchâtel, BAS at Basel and SMA at Zürich) have average temperatures just above 0°C during the winter, and between 15°C and 20°C during the summer. The other stations (JUN at Jungfraujoch, ANT at Andermatt and GLA at Glarus) are located in the Swiss Alps, and at higher elevations for stations JUN and ANT (respectively 3580 m and 1438 m). Temperatures are obviously lower for these stations. Station JUN is representative of highly mountainous areas. Station GLA is located in a valley and its average temperatures correspond more or less to the stations of the Swiss Plateau.

2.2 Multi-site temperature model

Let $T_t(k)$ denote the daily average temperature at location $k = 1, \ldots, K$ and time t, where $K = 26$ is the number of stations. This section presents a multi-site model reproducing several statistical properties of $T_t(k)$: long-term trends (non-stationarity), the annual cycle (seasonality), the statistical distribution modeling standardized temperature data at each station and inter-site dependence structures.

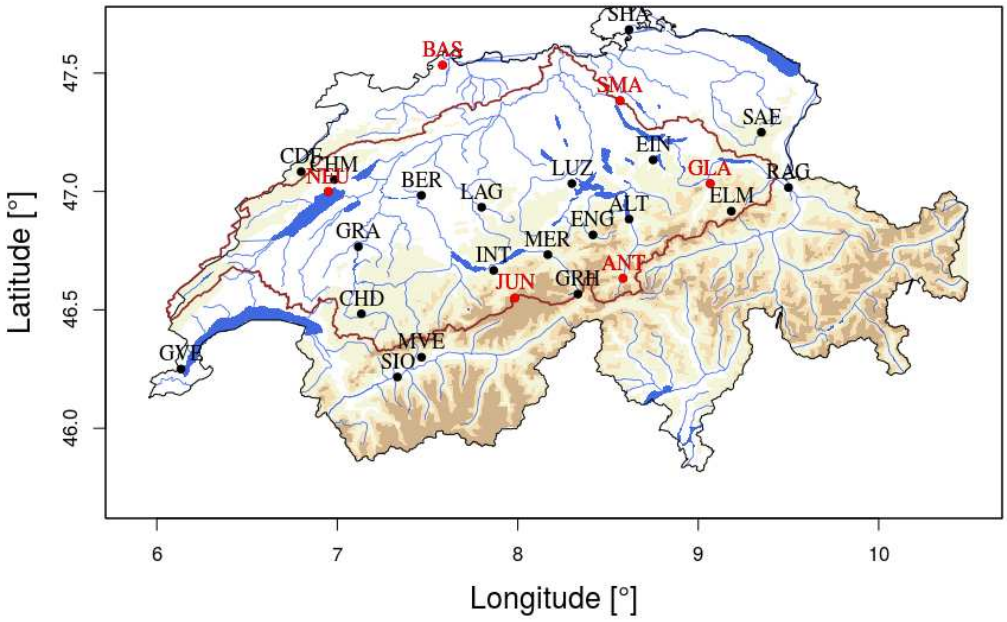


Figure 1: Location of the 26 temperature stations in Switzerland. The brown line indicates the limits of the Aare river catchment. A selection of 6 representative stations is highlighted in red.

Table 1: Description of 6 illustrative stations: Code, full name; elevation, winter temperature means (December, January and February), winter temperature minima, summer temperature means (June, July and August) and summer temperature maxima. The temperature statistics are in °C and correspond to the average metrics over all years (1930-2014).

Code	Full Name	Elevation [m]	Mean temp. in winter	Min. temp. in winter	Mean temp. in summer	Max. temp. in summer
NEU	Neuchâtel	485	1.4	-7.1	18.0	24.7
BAS	Basel / Binningen	316	1.8	-8.3	17.9	25.2
SMA	Zürich / Fluntern	555	0.4	-9.0	17.0	24.2
JUN	Jungfraujoch	3580	-13.5	-25.8	-1.4	5.4
ANT	Andermatt	1438	-4.7	-16.9	11.7	18.7
GLA	Glarus	516	-0.4	-10.6	16.6	23.4

2.2.1 Non-stationarity

Let $\bar{T}_m(y)$ denote the average regional temperature for the month m and the year y, i.e. the average temperature for all the stations. Figure 2 shows the time series of $\bar{T}_m(y)$ for the month of July. These regional temperatures clearly increase as a function of time, even if there is a high variability of the signal from one year to another (for example the very hot European summer in 2003 clearly stands out as atypical). This increase is directly related to the global warming identified by the scientific community (IPCC, 2014) and is thus important to consider in our model.

In this work, we propose to model this increase by a simple linear model:

$$\bar{T}_m(y) = \alpha + \beta \times y + \epsilon. \tag{1}$$

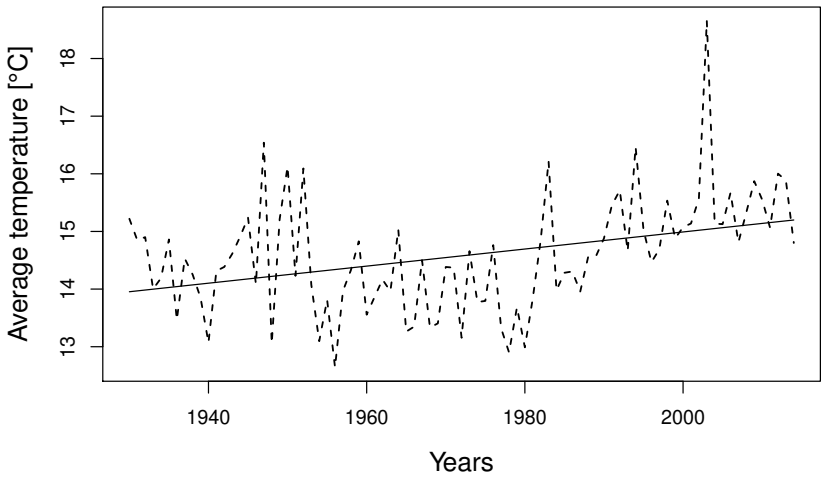


Figure 2: Time series of average regional temperatures for the month of July (dashed line) and fitted linear model (plain line).

This linear model is applied for each month and parameters α and β are estimated using the linear least squares method (see Fig. 2 for an illustration). Figure 3 shows the parameter estimates $\hat{\beta}$, which indicate the temperature increases per year, as a function of the month. These estimates are all significantly different from 0 at the 0.05 level. The increase of temperature is clearly higher for winter

months than for summer months, which means that the difference between winter and summer temperatures tends to decrease.

Differences between estimates $\hat{\beta}$ are large from one month to another. The jumps are most likely due to the uncertainty related to the parameter estimation. In order to obtain a more regular annual cycle of these estimates, a local polynomial model (loess) is applied to $\hat{\beta}$ (see Fig. 3).

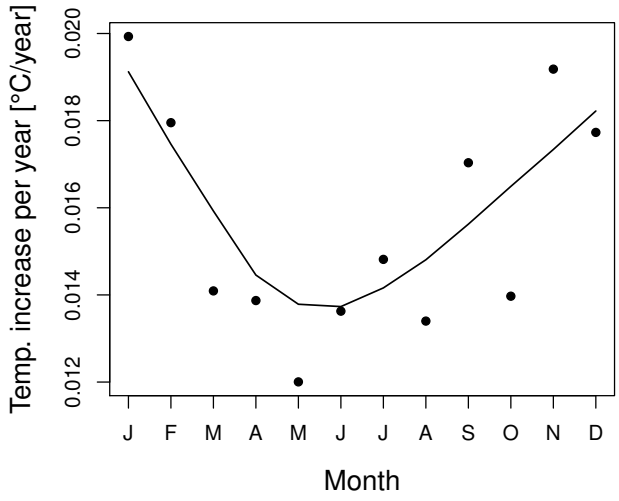


Figure 3: Estimates temperature increase per year ($\hat{\beta}$) for each month (dots) and fitted local polynomial model (plain line).

Linear trends are then removed from temperature data $T_t(k)$ using the smooth estimates $\hat{\alpha}$ and $\hat{\beta}$. These "stationary" temperature data are denoted by $\tilde{T}_t(k)$.

2.2.2 Seasonality

Annual cycles of temperature are often modeled by periodic functions, with sine and cosine curves (i.e. using a Fourier transform, see, e.g., Erhardt et al., 2015). In this model, for each location k, we use a more direct representation of the annual cycle:

- **daily mean temperature**: For each calendar day $d = 1, \ldots, 365$ of the year, we compute the mean of the "stationary" temperature data $\tilde{T}_t(k)$ corresponding to this day d. A loess smoothing is then applied to the average temperature in order to obtain the seasonal cycle of the daily mean temperature $\mu_d^{\tilde{T}}(k)$.

- **standard deviation of daily mean temperature**: Similarly, we compute standard deviations of the daily "stationary" temperature data for each day d, which are smoothed using a loess regression. This seasonal cycle of standard deviation of daily mean temperature is denoted by $\sigma_d^{\tilde{T}}(k)$.

Figure 4 shows the seasonal cycles of $\mu_d^{\tilde{T}}(k)$ (left plot) and $\sigma_d^{\tilde{T}}(k)$ (right plot) for the 6 illustrative stations. Concerning the mean temperature $\mu_d^{\tilde{T}}(k)$, the lower temperatures at stations ANT and JUN clearly stand out. For the mean temperature, we compare the seasonal cycles obtained using a non-parametric method (plain lines) and a periodic function, i.e. using a Fourier transform (dash lines). The non-parametric method and periodic functions lead to similar results. The seasonal cycles of the standard deviation $\sigma_d^{\tilde{T}}(k)$ indicate that the dispersion of temperatures are more important during winter than summer. For station NEU, the minimum variability occurs during autumn (Sep.-Nov.).

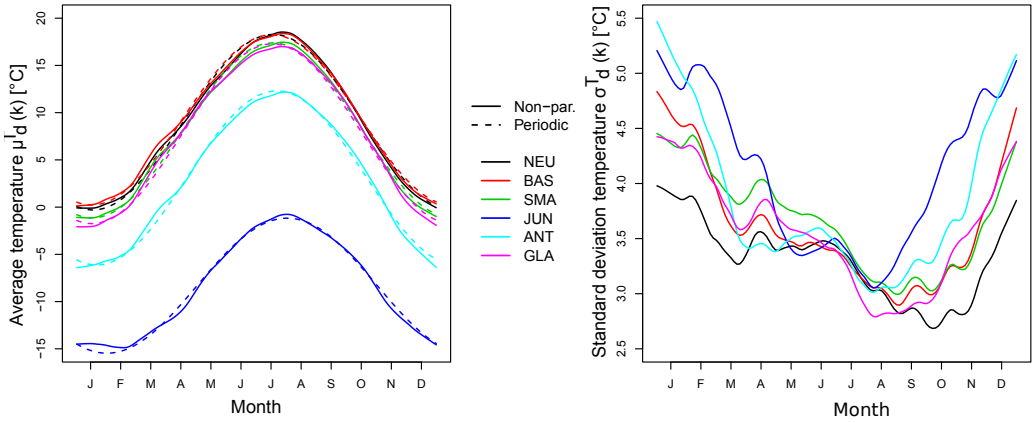


Figure 4: Seasonal cycles for the mean $\mu_d^{\tilde{T}}(k)$ (left) and for the standard deviation $\sigma_d^{\tilde{T}}(k)$ (right) of the "stationary" temperature data ($\tilde{T}_t(k)$), for the 6 illustrative stations.

These annual cycles are then repeated to match the period covered by the observations. For each location, these time series of annual cycles for the daily mean and standard deviation of temperature data are denoted by $\mu_t^{\tilde{T}}(k)$ and $\sigma_t^{\tilde{T}}(k)$, respectively. Standardized time series of temperature are then obtained as:

$$Z_t(k) = \{\tilde{T}_t(k) - \mu_t^{\tilde{T}}(k)\}/\sigma_t^{\tilde{T}}(k). \quad (2)$$

2.2.3 Marginal distributions

For each station k, the Skew Exponential Power (SEP) distribution, also known as the skew generalized error distribution (Fernandez and Steel, 1998), is fitted to standardized temperature data $Z_t(k)$, for each month. The SEP distribution is a generalization of the normal distribution and is defined in Appendix A. Very different shapes can be obtained according to the values taken by two additional parameters ξ and ν:

1. The parameter $\xi > 0$ introduces different levels of asymmetry (skewness). When $\xi = 1$, the distribution is symmetric and corresponds to the Gaussian case. When $\xi < 1$ (respectively $\xi > 1$), the distribution is skewed to the left (respectively to the right).
2. The parameter $\nu > 0$ is related to the flatness (kurtosis). When $\nu = 2$, the distribution corresponds to the Gaussian case. When $\nu < 2$ (respectively $\nu > 2$), the distribution has heavier (respectively lighter) tails than the normal distribution.

This distribution has been applied in various fields for its flexibility, for example in hydrology for residual errors, i.e. differences between observed and simulated streamflows (Schoups and Vrugt, 2010). Due to the standardization process (2), the mean and standard deviation of standardized temperature data can be considered to be equal to 0 and 1, respectively (see Eq. 2). Parameters ξ and ν of the SEP distribution are estimated monthly using the maximum likelihood method.

Figure 5 shows four examples of fitted SEP distributions and illustrates how this distribution can strongly depart from the Gaussian distribution depending on the month and on the station. Figure 6 shows the parameter estimates of ξ (left) and ν (right) for each month and for the 6 illustrative stations. Standardized temperatures $Z_t(k)$ are left-skewed in winter, particularly at high elevations (stations ANT and JUN). Obtained tails are particularly light during spring and summer months, at the exception of station JUN.

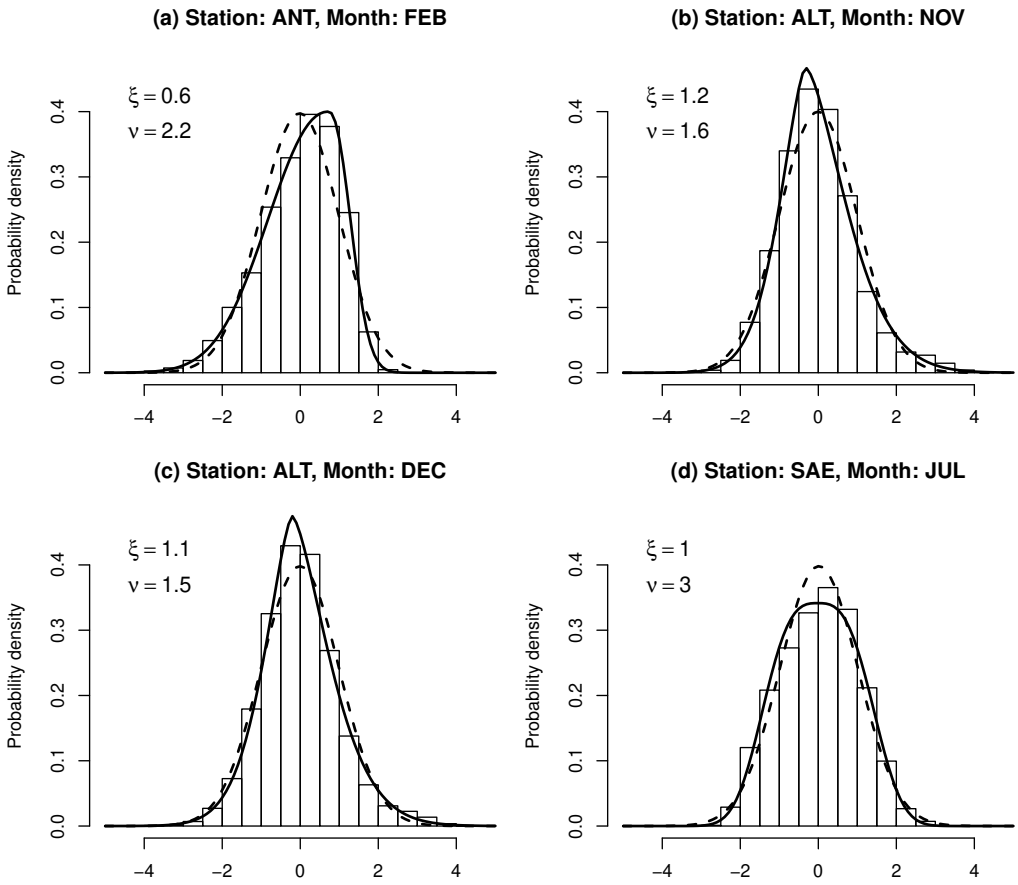


Figure 5: Examples of SEP densities (plain line) and Gaussian densities (dashed line) fitted to standardized temperatures $Z_t(k)$, for different stations and months. (a) Left-skewed. (b) Right-skewed. (c) Heavy tails. (d) Light tails.

2.2.4 Inter-site dependence

For each station, the probability integral transform is first applied to standardized temperatures $Z_t(k)$ in order to remove the marginal effects:

$$U_t(k) = \big[F_Z\{Z_t(k)|\nu, \xi\}\big], \tag{3}$$

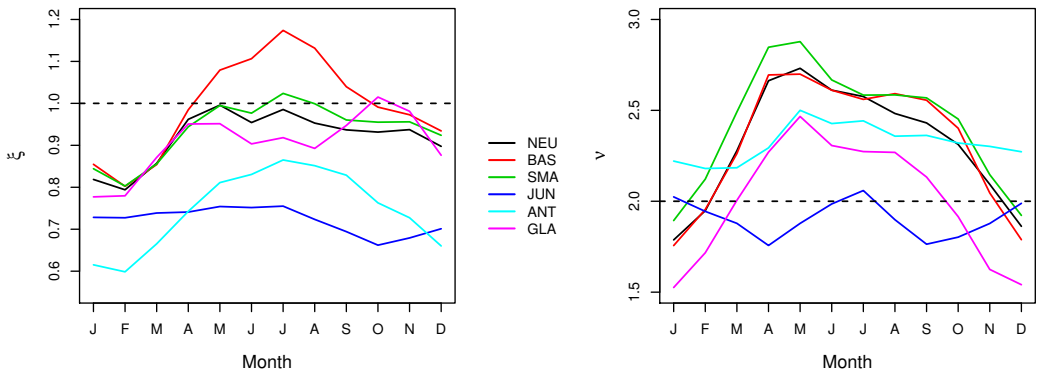


Figure 6: Parameter estimates of ξ (left) and ν (right) for the 6 illustrative stations.

where F_Z denotes the cumulative distribution function of the marginal distribution, i.e. the SEP distribution. Random variables $U_t(k)$ thus follow uniform distributions.

The spatial dependence of temperature is usually represented using multivariate Gaussian distributions (see, e.g. Baigorria, 2014; Chen et al., 2014). "Gaussianized" temperature variates are defined as

$$X_t(k) = \Phi^{-1}\{U_t(k)\}, \tag{4}$$

where Φ denotes the cumulative distribution function of the Gaussian distribution. $\mathbf{X}_t = \{X_t(1), \ldots, X_t(K)\}$ is assumed to be a vector of K Gaussian variates with mean 0.

In this study, a Multivariate Autoregressive model of order 1, $MAR(1)$ (Bárdossy and Pegram, 2009; Rasmussen, 2013) represents simultaneously spatial and temporal inter-sites dependences. The $MAR(1)$ process can be applied to "gaussianized" temperature variates (4) as follows:

$$\mathbf{X}_t = \mathbf{A}\mathbf{X}_{t-1} + \epsilon_t, \tag{5}$$

where $\mathbf{A}$ is a $K \times K$ matrix and ϵ_t is a random $K \times 1$ noise vector. Innovations ϵ_t are assumed to follow a standard multivariate normal distribution, its elements being independent of the elements of $\mathbf{Z}_{t-1}$. The covariance matrix of ϵ_t is denoted by Ω_X. Matalas (1967) shows

that $\mathbf{A}$ and Ω_X can be expressed in terms of the covariance matrices of $\mathbf{X}_t$:

$$\mathbf{A} = \mathbf{M}_1 \mathbf{M}_0^{-1}, \tag{6}$$

$$\Omega_X = \mathbf{M}_0 - \mathbf{M}_1 \mathbf{M}_0^{-1} \mathbf{M}_1', \tag{7}$$

where $\mathbf{M}_0$ and $\mathbf{M}_1$ are the covariance matrices of $\mathbf{X}_t$ at lag-0 and lag-1, respectively. The matrices $\mathbf{M}_0$ and $\mathbf{M}_1$ are estimated by the corresponding cross-correlations for all possible pairs of stations using the Pearson correlation coefficient ρ. Since $\mathbf{M}_0$ and $\mathbf{M}_1$ are not necessarily positive-definite, they are replaced by the closest positive-definite matrices if necessary (Rousseeuw and Molenberghs, 1993; Rebonato and Jaeckel, 2011). Similarly than for the marginal distributions, the seasonality of the multivariate dependence is taken into account by estimating $\mathbf{A}$ and Ω_X for each month.

3 Case study

3.1 Model versions

In this section, we evaluate the performances of three different versions of the temperature model described above:

1. **GAU**: The first model has Gaussian marginal distributions at each station and a Gaussian spatial dependence. The temporal dependence is ignored, i.e. $\mathbf{A}$ is a zero matrix in Eq. (5). This version can be considered as a benchmark, most stochastic temperature models adopting this structure (see, e.g. Baigorria, 2014; Chen et al., 2014).
2. **SEP**: The second version is a modification of the model GAU where the SEP distribution is the marginal distribution applied at each station.
3. **MAR1**: The last version evaluated in this study applies the SEP distribution to margins and a $MAR(1)$ process (see Eq. 5) to model inter-site spatial and temporal dependences.

The characteristics of these three different versions are summarized in Table 2. The impact of the different statistical components can thus be assessed with cross-comparisons of these models.

Table 2: Summary of the characteristics of the temperature models.

Model name	Marginal distribution	Spatial dependence	Temporal dependence
GAU	Gaussian	✓	X
SEP	SEP	✓	X
MAR1	SEP	✓	✓

For each of these three temperature models, 100 synthetic temperature scenarios are compared to the observed temperature series at the 26 stations, each scenario having the same length as the observed series (i.e. 85 years). These scenarios are generated under a stationary climate and are compared to the "stationary" temperature data $\tilde{T}_t(k)$ (see Section 2.2.1).

3.2 Inter-site correlations

Figure 7 confronts observed and synthetic inter-site unlagged and lag-1 correlations for the winter and summer seasons and for the different model versions. Unlagged correlations, which represent the spatial dependence, are close to the 1:1 line, which is expected as these correlations are explicitly taken into account by all model versions. Inter-site correlations are higher during the summer than during the winter. Since spatial correlations are adequately reproduced by this stochastic generator, we can assume that realistic daily temperature fields are simulated, which can be important in many applications. For example, in hydrological applications, areal temperature estimates are used to force conceptual hydrological models. Areal temperature data impacts, among other things, the proportion of precipitation falling as snow and rainfall in a catchment, and snowmelt runoff, which can be of crucial importance for the reproduction of flood characteristics (peak, duration, etc.).

An adequate reproduction of crossed lag-1 correlations is important for the reproduction of persistent temperature properties, such as cold spells and heat waves (see below). Simulated lag-1 correlations are close to 0 for models GAU and SEP, which is expected since these versions ignore the temporal dependence. Consequently, the model versions significantly underestimate observed lag-1 correlations, which range between 0.2 and 0.8 in the winter and between 0.6 and 0.8 in the summer. This under-estimation is not present for MAR1, thanks to the MAR(1) process (see Section 2.2.4).

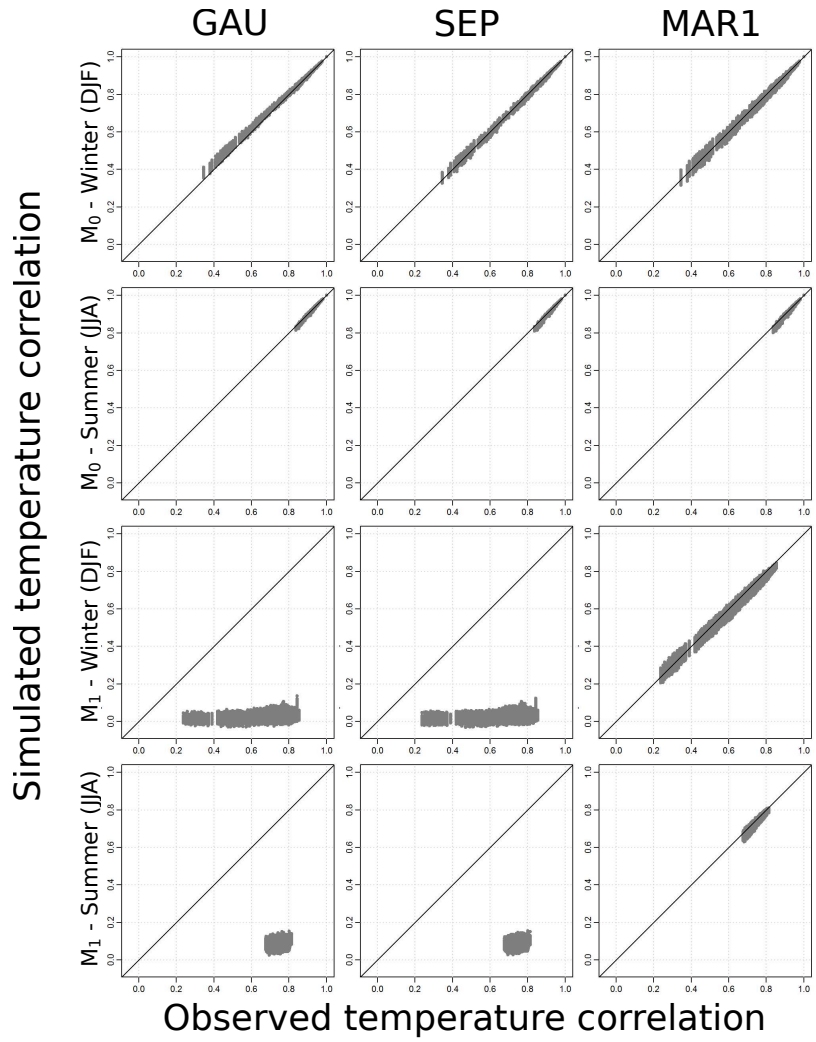


Figure 7: Comparison of unlagged ($\mathbf{M}_0$) and lagged ($\mathbf{M}_1$) correlations in observed and synthetic temperature series, for the winter and summer seasons, and for the different model versions considered.

3.3 Temperature lapse rate

Figure 8 represents the probability densities of the temperature gradients between stations BAS and JUN for the winter and summer seasons and for the different model versions. In this study, stations BAS and JUN are the stations corresponding to the lowest (316 m)

and highest (3580 m) elevations, respectively. The difference of temperatures, expressed as a temperature lapse rate in °C/100m (temperature decrease with an increase of 100m in altitude) between these two stations is an interesting indicator of the elevation-temperature relationship, which can vary from day to day according to the meteorological situation.

The probability densities of observed and synthetic lapse rates are in good agreement for all model versions, for both winter and summer seasons. The temperature lapse rates vary in a wide range during winter (between 0°C/100m and 0.8°C/100m), while these lapse rates are concentrated between 0.4°C/100m and 0.75°C/100m during summer. Low temperature lapse rates are thus more frequent during the winter. Such temperature lapse rates are unusual and can be interpreted as "temperature inversions", which a well-known phenomenon. Under this meteorological conditions, there is very few exchanges of air between the Swiss Plateau and the rest of the atmosphere, cold air being trapped at low elevations. Figure 8 demonstrates that the stochastic generators presented in this study provide a reasonable representation of this meteorological configuration, in terms of frequency, thanks to an adequate and flexible structure of spatial dependence.

3.4 Temperature extremes

Extreme temperature events affect the physiology, development, ecology and evolution of plants (Gutschick and BassiriRad, 2003). For example, extreme temperatures impact plant productivity, especially after the pollination (Hatfield and Prueger, 2015). They can also lead to early bud breaks (Luedeling, 2012) or late frosts (Augspurger, 2013), which are critical phenomena in different branches of agriculture (forestry, horticulture). Here, we focus on the analysis on winter maxima and spring minima, assuming that winter maxima can be related to the chances of breaking winter dormancy, and that spring minima is a good indicator of late frosts. Figure 9 compares observed and simulated extremes of daily temperature associated to a return period of $T = 20$ years, for the different model versions. Observed extremes are clearly underestimated by model GAU. This underestimation is greatly reduced for model SEP, which indicates that the SEP distribution provides a more adequate fit of the marginal distribution tails (see Section 2.2.3). Model MAR1 seems to yield to a better agreement between observed and simulated spring minima.

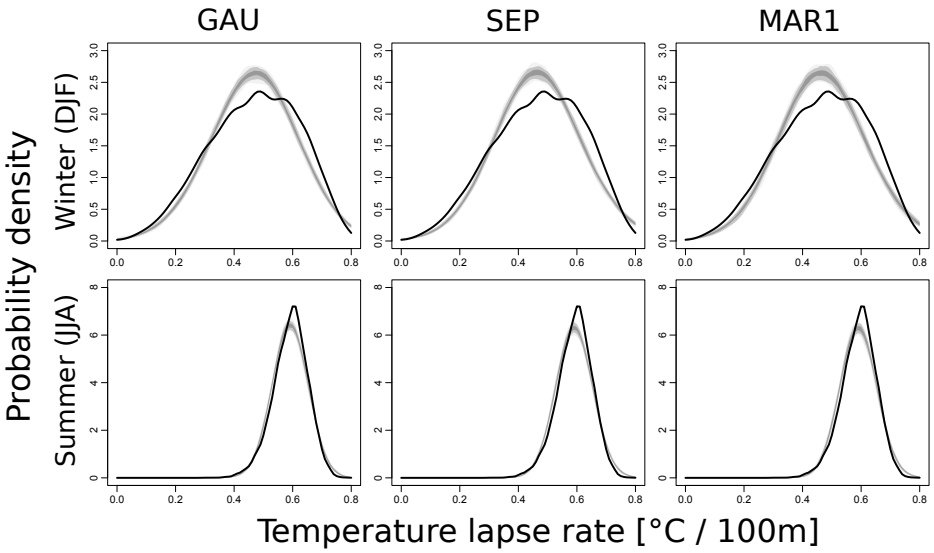


Figure 8: Probability densities of the temperature gradients between stations BAS and JUN, for the winter and summer seasons, and the different model versions considered. Observed probability densities (in black) can be compared to the 99% confidence intervals of simulated probability densities (in gray).

The temporal autocorrelation also plays a role in the reproduction of temperature extremes, which could be explained by the persistence of extreme temperature events (e.g. cold spells and heat waves).

3.5 Heat waves

Heat waves directly impact the economic activity (e.g. crop failures), transport infrastructure (buckling of rail tracks, damage to roads), electricity demand and human health (Zuo et al., 2015). A striking example is the 2003 European heat wave which has been directly responsible of more than 70,000 deaths in Europe (Robine et al., 2008). In this study, heat waves are detected according to the methodology proposed by Ouzeau et al. (2016) and are defined as periods of consecutive days with a daily average temperature above the 97.5% quantile. The 97.5% quantiles are computed at each station using observed "stationary" temperature data $\tilde{T}_t(k)$. To be recorded as an heat wave, this period must exceed 2 days and include a temperature peak above the

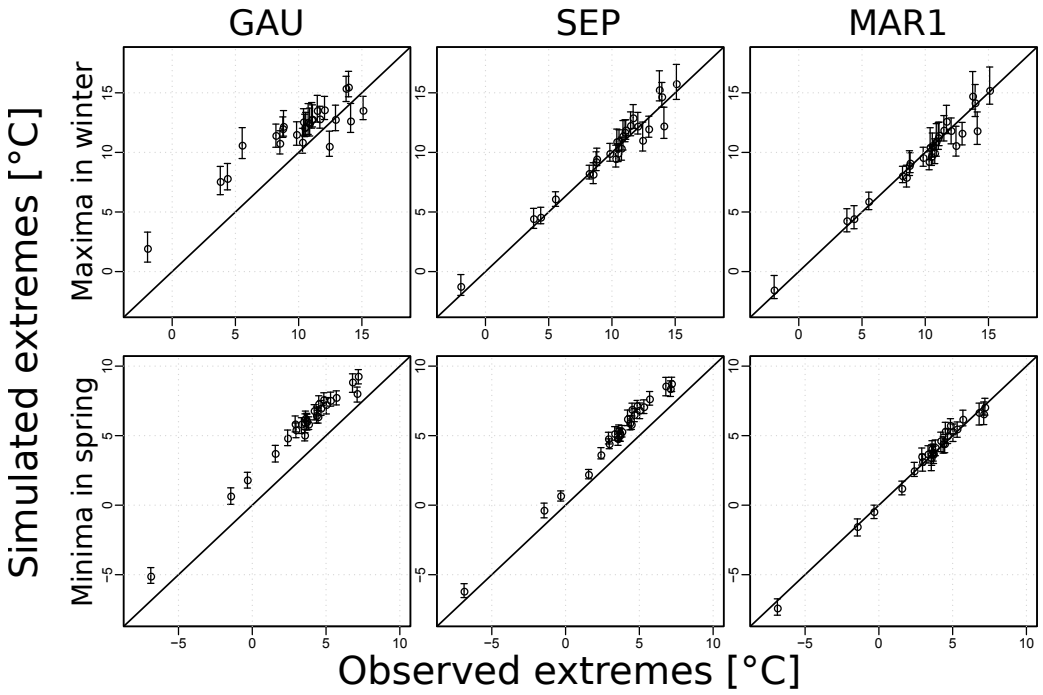


Figure 9: Observed and simulated extremes of daily temperature associated to a return period of $T = 20$ years: maxima during the winter season (DJF, top plots) and minima during the spring season (MAM, bottom plots), for the different models considered. The confidence intervals indicate 95% probability limits.

99.5% quantile. Figure 10 presents the frequencies of observed and simulated heat waves, according to their lengths, for three illustrative stations and for the different models considered. Models GAU and SEP rarely generate heat waves, and, as a consequence, underestimate dramatically the frequencies of observed heat waves. This is not the case for model MAR1, for which the MAR(1) process (see Eq. 5) provides an adequate representation of the persistence of high daily temperatures.

3.6 Cold spells

Cold spells are also associated with increased mortality rates (see, e.g., Analitis et al., 2008, for a study in Europe). Moreover, extreme

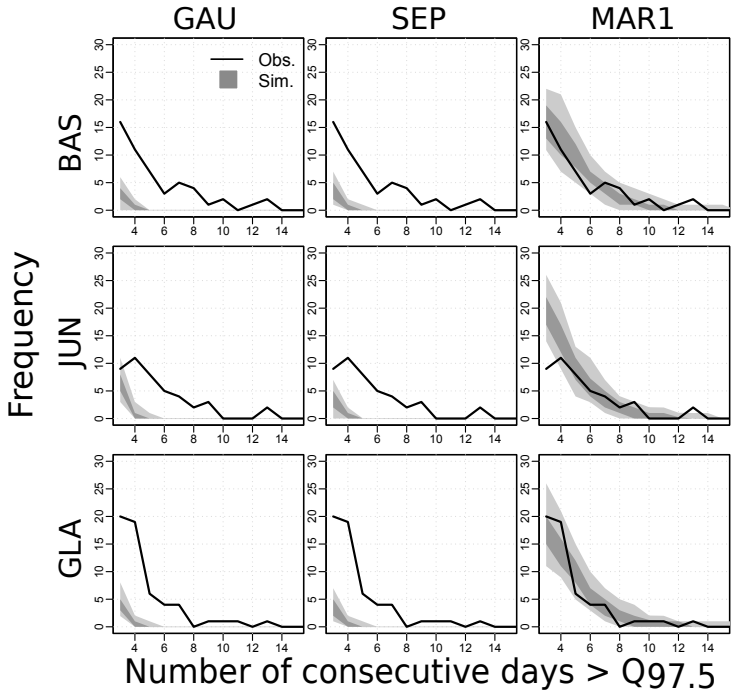


Figure 10: Frequency of simulated and observed heat waves, as a function of their lengths, for three illustrative stations and for the different models considered: 50% and 90% probability limits are shown. Heat waves are selected as the periods of consecutive days with a daily average temperature above the 97.5% quantile.

winter cold cause significant damages to infrastructures (gas and water pipeline disruptions), as well as a dramatic increase of the demand for electrical power during such these periods.

In this study, cold spells are defined as periods of consecutive days with a daily average temperature below the 2.5% quantile, computed at each station using observed temperature. Cold spells are retained if their duration exceeds 4 days and include a temperature peak below the 0.5% quantile. Figure 11 presents the frequencies of observed and simulated cold spells, according to their lengths, for three illustrative stations and for the different models considered. Models GAU and SEP almost never generate persistent cold spells. Long cold spells can only be simulated using model MAR1, which explicitly takes into account the persistence of low daily temperatures.

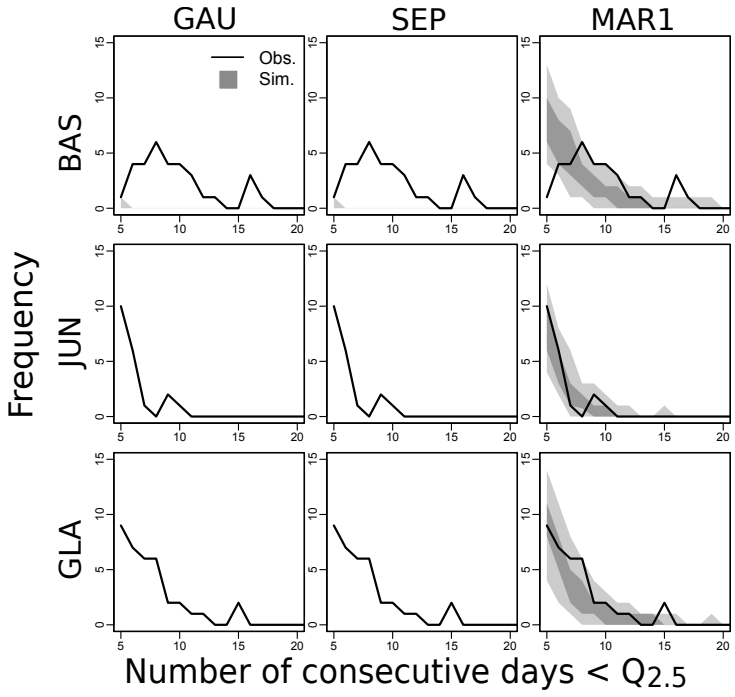


Figure 11: Frequency of simulated and observed cold spells, as a function of their lengths, for three illustrative stations and for the different models considered: 50% and 90% probability limits are shown. Cold spells are selected as the periods of consecutive days with a daily average temperature below the 2.5% quantile.

4 Conclusions and outlook

This paper presents a flexible and robust multi-site daily temperature model which can be applied to a high number of stations and various climates. The main novelties of this temperature model are:

- the representation of seasonal cycles with non-parametric functions,
- the application of the SEP distribution to the standardized temperature at each station. This 4-parameter distribution generalizes the Gaussian distribution and can reproduce different skewness and kurtosis,
- the introduction of a Multivariate Autoregressive (MAR) model to represent the spatial and temporal dependence structures simultaneously.

This study demonstrates how these statistical features are important components of stochastic temperature generators. The SEP distribution leads to a better representation of daily temperature extremes (e.g. winter maxima, spring minima). Inter-site lagged correlations must be explicitly taken into account in order to reproduce persistent temperature phenomena (e.g. heat waves and cold spells). Depending on the application (agriculture, forestry, hydrology), we discuss the importance of these statistical components in regard to the reproduction of specific temperature events.

Future developments will be devoted to the introduction of a link between synthetic precipitation and temperature series. Indeed, the relationship between precipitation and temperature can be of first importance. For example, in the context of flood risk assessment, how temperature impacts the phase of precipitation falls (proportion of precipitation falling as snow or rain) has direct consequences on floods. Different strategies can be considered to introduce such a link:

- The temperature model can be parameterized differently according to precipitation states (e.g. dry and wet states, see Wilks, 2009).
- Atmospheric predictors can be used to drive the different meteorological variables (Mezghani and Hingray, 2009; Asong et al., 2016; Chandler, 2016).

Furthermore, for some applications (for example in forestry), fine spatial resolutions are needed (e.g. on a 30m x 30m grid). The multi-

site temperature generators presented in this study could be combined to interpolators dedicated to temperature fields (Frei, 2014).

5 Acknowledgments

Financial support for this study by the Swiss Federal Office for Environment (FOEN), the Swiss Federal Nuclear Safety Inspectorate (ENSI), the Federal Office for Civil Protection (FOCP) and the Federal Office of Meteorology and Climatology, MeteoSwiss, through the project EXAR ("Evaluation of extreme Flooding Events within the Aare-Rhine hydrological system in Switzerland"), is gratefully acknowledged.

A Skew Exponential distribution

The Exponential Power distribution, also known as the generalized error distribution, is a generalization of the normal distribution. A shape parameter, $\nu > 0$, leads to tails that are either heavier than normal if $\nu < 2$ and lighter if $\nu > 2$, $\nu = 2$ corresponds to the normal case. For a standardized variate z (with a mean of 0 and a standard deviation of 1), its probability density is defined as:

$$f_{EP}(z|\nu) = \frac{\nu}{\lambda_\nu \times 2^{1+1/\nu} \times \Gamma(1/\nu)} \times \exp\Big(-\frac{1}{2} \times \Big|\frac{z}{\lambda_\nu}\Big|^\nu \Big),$$

with

$$\lambda_\nu = \sqrt{2^{-2/\nu} \times \frac{\Gamma(1/\nu)}{\Gamma(3/\nu)}}.$$

The Skew Exponential Power (SEP) distribution is a generalization of the Skew Exponential distribution. An additional shape parameter, ξ, can lead to different skewness when it differs from 1. For a normalized variate z and parameters $\nu > 0$ and $\xi > 0$, the probability density of the SEP distribution is defined as:

$$f_{SEP}(z|\nu,\xi) = \frac{2}{\xi + 1/\xi}\, f_{EP}(\tilde{z}/\xi^{\mathrm{sign}(\tilde{z})}|\nu), \tag{8}$$

with

$$\begin{aligned}
\tilde{z} &= \mu_z + \sigma_z \times z, \\
\mu_z &= \gamma_\nu \times (\xi - 1/\xi), \\
\sigma_z &= \sqrt{(1-\gamma_\nu^2)(\xi^2 + 1/\xi^2) + 2\gamma_\nu^2 - 1}, \\
\gamma_\nu &= 2^{1/\nu} \times \lambda_\nu \times \frac{\Gamma(2/\nu)}{\Gamma(1/\nu)}.
\end{aligned}$$

References

Analitis, A., Katsouyanni, K., Biggeri, A., Baccini, M., Forsberg, B., Bisanti, L., Kirchmayer, U., Ballester, F., Cadum, E., Goodman,

P. G., Hojs, A., Sunyer, J., Tiittanen, P., and Michelozzi, P. (2008). Effects of Cold Weather on Mortality: Results From 15 European Cities Within the PHEWE Project. *American Journal of Epidemiology*, 168(12):1397–1408.

Apipattanavis, S., Podestá, G., Rajagopalan, B., and Katz, R. W. (2007). A semiparametric multivariate and multisite weather generator. *Water Resources Research*, 43(11):W11401.

Asong, Z. E., Khaliq, M. N., and Wheater, H. S. (2016). Multisite multivariate modeling of daily precipitation and temperature in the Canadian Prairie Provinces using generalized linear models. *Climate Dynamics*, 47(9-10):2901–2921.

Augspurger, C. K. (2013). Reconstructing patterns of temperature, phenology, and frost damage over 124 years: Spring damage risk is increasing. *Ecology*, 94(1):41–50.

Baigorria, G. A. (2014). Stochastic Models to Generate Geospatial-, Temporal-, and Cross-Correlated Daily Maximum and Minimum Temperatures. *Advances in Meteorology*. Vol. 2014, Article ID 365362, 14 pages.

Bárdossy, A. and Pegram, G. G. S. (2009). Copula based multisite model for daily precipitation simulation. *Hydrology and Earth System Sciences*, 13(12):2299–2314.

Boumans, R. J. M., Phillips, D. L., Victery, W., and Fontaine, T. D. (2014). Developing a model for effects of climate change on human health and health–environment interactions: Heat stress in Austin, Texas. *Urban Climate*, 8:78–99.

Breinl, K., Turkington, T., and Stowasser, M. (2015). Simulating daily precipitation and temperature: a weather generation framework for assessing hydrometeorological hazards. *Meteorological Applications*, 22(3):334–347.

Chandler, R. E. (2016). Rglimclim: A multisite, multivariate weather generator based on generalized linear models. R package version 1.3-0. http://www.ucl.ac.uk/~ucakarc/work/glimclim.html.

Chandler, R. E. and Wheater, H. S. (2002). Analysis of rainfall variability using generalized linear models: A case study from the west of Ireland. *Water Resources Research*, 38(10):1192.

Chen, J., Brissette, F. P., and Zhang, J. X. (2014). A Multi-Site Stochastic Weather Generator for Daily Precipitation and Temperature. *Transactions of the ASABE*, pages 1375–1391.

Deryng, D., Conway, D., Ramankutty, N., Price, J., and Warren, R. (2014). Global crop yield response to extreme heat stress under multiple climate change futures. *Environmental Research Letters*, 9(3):034011.

Erhardt, T. M., Czado, C., and Schepsmeier, U. (2015). R-vine models for spatial time series with an application to daily mean temperature. *Biometrics*, 71(2):323–332.

Fernandez, C. and Steel, M. F. J. (1998). On Bayesian Modeling of Fat Tails and Skewness. *Journal of the American Statistical Association*, 93(441):359–371.

Frei, C. (2014). Interpolation of temperature in a mountainous region using nonlinear profiles and non-Euclidean distances. *International Journal of Climatology*, 34(5):1585–1605.

Guenet, B., Cadule, P., Zaehle, S., Piao, S. L., Peylin, P., Maignan, F., Ciais, P., and Friedlingstein, P. (2013). Does the integration of the dynamic nitrogen cycle in a terrestrial biosphere model improve the long-term trend of the leaf area index? *Climate Dynamics*, 40(9-10):2535–2548.

Gutschick, V. P. and BassiriRad, H. (2003). Extreme events as shaping physiology, ecology, and evolution of plants: toward a unified definition and evaluation of their consequences. *New Phytologist*, 160(1):21–42.

Hatfield, J. L. and Prueger, J. H. (2015). Temperature extremes: Effect on plant growth and development. *Weather and Climate Extremes*, 10:4–10.

IPCC (2014). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report*

of the Intergovernmental Panel on Climate Change. Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.), IPCC, Geneva, Switzerland. 151pp.

King, L. M., McLeod, A. I., and Simonovic, S. P. (2014). Simulation of historical temperatures using a multi-site, multivariate block resampling algorithm with perturbation. *Hydrological Processes*, 28(3):905–912.

King, L. M., McLeod, A. I., and Simonovic, S. P. (2015). Improved Weather Generator Algorithm for Multisite Simulation of Precipitation and Temperature. *JAWRA Journal of the American Water Resources Association*, 51(5):1305–1320.

Krinner, G., Viovy, N., de Noblet-Ducoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., Ciais, P., Sitch, S., and Prentice, I. C. (2005). A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system. *Global Biogeochemical Cycles*, 19(1). GB1015.

Kucharik, C. J., Foley, J. A., Delire, C., Fisher, V. A., Coe, M. T., Lenters, J. D., Young-Molling, C., Ramankutty, N., Norman, J. M., and Gower, S. T. (2000). Testing the performance of a dynamic global ecosystem model: Water balance, carbon balance, and vegetation structure. *Global Biogeochemical Cycles*, 14(3):795–825.

Leander, R. and Buishand, T. A. (2009). A daily weather generator based on a two-stage resampling algorithm. *Journal of Hydrology*, 374(3–4):185–195.

Li, Z. (2014). A new framework for multi-site weather generator: a two-stage model combining a parametric method with a distribution-free shuffle procedure. *Climate Dynamics*, 43(3-4):657–669.

Luedeling, E. (2012). Climate change impacts on winter chill for temperate fruit and nut production: A review. *Scientia Horticulturae*, 144:218–229.

Matalas, N. C. (1967). Mathematical assessment of synthetic hydrology. *Water Resources Research*, 3(4):937–945.

Mezghani, A. and Hingray, B. (2009). A combined downscaling-disaggregation weather generator for stochastic generation of multisite hourly weather variables over complex terrain: Development and multi-scale validation for the Upper Rhone River basin. *Journal of Hydrology*, 377(3–4):245–260.

Ouzeau, G., Soubeyroux, J. M., Schneider, M., Vautard, R., and Planton, S. (2016). Heat waves analysis over France in present and future climate: Application of a new method on the EURO-CORDEX ensemble. *Climate Services*, 4:1–12.

Rasmussen, P. F. (2013). Multisite precipitation generation using a latent autoregressive model. *Water Resources Research*, 49(4):1845–1857.

Rebonato, R. and Jaeckel, P. (2011). The Most General Methodology to Create a Valid Correlation Matrix for Risk Management and Option Pricing Purposes. SSRN Scholarly Paper ID 1969689, Social Science Research Network, Rochester, NY.

Richardson, C. W. (1981). Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research*, 17(1):182–190.

Robine, J.-M., Cheung, S. L. K., Le Roy, S., Van Oyen, H., Griffiths, C., Michel, J.-P., and Herrmann, F. R. (2008). Death toll exceeded 70,000 in Europe during the summer of 2003. *Comptes Rendus Biologies*, 331(2):171–178.

Romero, C. C., Dukes, M. D., Baigorria, G. A., and Cohen, R. (2009). Comparing theoretical irrigation requirement and actual irrigation for citrus in Florida. *Agricultural Water Management*, 96(3):473–483.

Rousseeuw, P. J. and Molenberghs, G. (1993). Transformation of non positive semidefinite correlation matrices. *Communications in Statistics - Theory and Methods*, 22(4):965–984.

Schoups, G. and Vrugt, J. A. (2010). A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. *Water Resources Research*, 46(10):W10531.

Steinschneider, S. and Brown, C. (2013). A semiparametric multivariate, multisite weather generator with low-frequency variability for use in climate risk assessments. *Water Resources Research*, 49(11):7205–7220.

Stöckle, C. O., Donatelli, M., and Nelson, R. (2003). CropSyst, a cropping systems simulation model. *European Journal of Agronomy*, 18(3–4):289–307.

Wilks, D. S. (1992). Adapting stochastic weather generation algorithms for climate change studies. *Climatic Change*, 22(1):67–84.

Wilks, D. S. (1998). Multisite generalization of a daily stochastic precipitation generation model. *Journal of Hydrology*, 210(1–4):178–191.

Wilks, D. S. (1999). Simultaneous stochastic simulation of daily precipitation, temperature and solar radiation at multiple sites in complex terrain. *Agricultural and Forest Meteorology*, 96(1–3):85–101.

Wilks, D. S. (2009). A gridded multisite weather generator and synchronization to observed weather data. *Water Resources Research*, 45(10):W10419.

Wilks, D. S. and Wilby, R. L. (1999). The weather generation game: a review of stochastic weather models. *Progress in Physical Geography*, 23(3):329–357.

Yang, Y., Liu, D. L., Anwar, M. R., Zuo, H., and Yang, Y. (2014). Impact of future climate change on wheat production in relation to plant-available water capacity in a semiarid environment. *Theoretical and Applied Climatology*, 115(3-4):391–410.

Zuo, J., Pullen, S., Palmer, J., Bennetts, H., Chileshe, N., and Ma, T. (2015). Impacts of heat waves and corresponding measures: a review. *Journal of Cleaner Production*, 92:1–12.

10.3.3 New family of Fisher copulas to model upper tail spatial dependence and radial asymmetry

Research Article **Environmetrics**

Received 00 Month 2017 Revised 00 Month 2017 Accepted 00 Month 2017

(wileyonlinelibrary.com) DOI: 10.1002/env.XXXX

The new family of Fisher copulas to model upper tail dependence and radial asymmetry: properties and application to high-dimensional rainfall data

Anne-Catherine Favre[a], Jean-François Quessy[b] and Marie-Hélène Toupin[c]*

Joint precipitation data measured at a large number of stations typically show tail asymmetry and significant upper tail dependence. Unfortunately, many multivariate dependence models that are commonly used in large dimensions like the Normal and the Student copulas are radially symmetric, while the recently introduced chi-square copula is asymmetric, but its tail dependence coefficients are null. In order to circumvent the limitations of the available models, the new family of Fisher copulas is introduced; it is shown that these dependence models are tail asymmetric and allow for upper tail dependence, among other characteristics. Two semi-parametric strategies for parameter estimation in this class of copulas are proposed and their efficiency in small and moderate sample sizes is investigated with the help of simulations. The usefulness of the parametric Fisher copula family is then illustrated on the modeling of precipitation data observed at 105 stations within or close to the Aare river catchment in Switzerland.

Keywords: High-dimensional models, pairwise likelihood, precipitation data, semi-parametric inference

1. INTRODUCTION

Long time series of precipitation and temperature are required as input for hydrological models. Since the samples of measured precipitation are often too small for hydrological purposes, synthetic series of precipitation generated by stochastic models are often required (see Sun and Stein (2015) and Lehmann *et al.* (2016), for instance). In the context of risk management, it is important to be able to reproduce joint precipitation extremes at several meteorological stations in order to be able to mimic the behavior of discharge extremes at several spatial scales (catchments or sub-catchments). A possible measure of the joint probability of occurrence of extreme events is provided by the upper tail dependence coefficient. Specifically, the upper tail dependence coefficient of a random pair (X_1, X_2) with marginal distributions F_1 and F_2 is the conditional probability

$$\lambda_U(X_1, X_2) = \lim_{u\uparrow 1} \mathbb{P}\left\{X_1 > F_1^{-1}(u) \middle| X_2 > F_2^{-1}(u)\right\}.$$

The random variables X_1 and X_2 are said to be asymptotically upper tail dependent if $\lambda_U(X_1, X_2) > 0$, and upper tail independent otherwise. A measure that is similar in nature to λ_U is the lower tail dependence coefficient, namely $\lambda_L(X_1, X_2) = \lim_{u\downarrow 0} \mathbb{P}\{X_1 \leq F_1^{-1}(u) | X_2 \leq F_2^{-1}(u)\}$.

Precipitation data generally consist of time series, for example daily observations measured at a large number of locations, such as meteorological stations. In that context, the use of multivariate copulas is very attractive: on one side, it allows for the construction of flexible models appearing as alternatives to traditional ones, and on the other side, dependence in these models can be interpreted independently of the marginal behaviors. Copula theory originates from Sklar's Theorem that states that for any random vector $\mathbf{X} = (X_1, \ldots, X_d)$ with marginal distributions $F_1, \ldots, F_d$, there exists a function $C : [0,1]^d \to [0,1]$ such that

$$\mathbb{P}(X_1 \leq x_1, \ldots, X_d \leq x_d) = C\left\{F_1(x_1), \ldots, F_d(x_d)\right\}.$$

[a] *Univ. Grenoble Alpes, CNRS, IRD, Grenoble INP, IGE, 38000 Grenoble, France. E-mail: anne-catherine.favre-pugin@ense3.grenoble-inp.fr*
[b] *Département de mathématiques et d'informatique, Université du Québec à Trois-Rivières, Canada. E-mail: jean-francois.quessy@uqtr.ca*
[c] *Département de mathématiques et de statistique, Université Laval, Québec, Canada. E-mail: marie-helene.toupin1@ulaval.ca*
* *Correspondence to: M.-H. Toupin, Département de mathématiques et statistique, Université Laval, Québec, Canada. E-mail: marie-helene.toupin1@ulaval.ca*

If $F_1, \ldots, F_d$ are continuous, C is unique and is called the copula of $\mathbf{X}$. In that case, C characterizes the dependence structure of $\mathbf{X}$ and corresponds to the joint distribution of $(U_1, \ldots, U_d)$, where $U_\ell = F_\ell(X_\ell)$. In particular, the lower and upper tail dependence coefficients of a continuous random pair (X_1, X_2) with copula C can be expressed as

$$\lambda_L(X_1, X_2) = \lambda_L(C) = \lim_{u \downarrow 0} \frac{C(u,u)}{u} \quad \text{and} \quad \lambda_U(X_1, X_2) = \lambda_U(C) = \lim_{u \downarrow 0} \frac{\bar{C}(u,u)}{u}, \tag{1}$$

where $\bar{C}(u,v) = u + v - 1 + C(1-u, 1-v)$ is the survival copula. Many other dependence measures are marginal-free in the sense that they can be defined in terms of C, independently of the marginal distributions. It is the case for the population version of Kendall's measure of association, initially defined as a difference between probabilities of concordance and discordance. In fact, the latter can be written as

$$\tau(X_1, X_2) = \tau(C) = 4 \int_0^1 \int_0^1 C(u_1, u_2) \, \mathrm{d}C(u_1, u_2) - 1.$$

The reader is referred to the monographs by Nelsen (2006), Mai and Scherer (2012) and Joe (2015) for further details on copulas and related notions of dependence.

However, the multivariate copula families that are currently available for high-dimensional modeling suffer from some limitations. For example, even though the Normal copula has well established theoretical properties and is easy to interpret, the kind of dependence that it can accommodate is limited to cases where the structure is radially symmetric and tail dependence vanishes. An alternative to the Normal is the Student copula which allows for non-null tail dependence, but like the normal, is radially symmetric; a similar comment applies to the elliptical class of copulas. The chi-square copulas introduced by Bàrdossy (2006) and investigated by Quessy *et al.* (2016) are radially asymmetric, however their tail dependence coefficients are null. Another possibility is to use the vine copulas popularized by Aas *et al.* (2009), but this approach necessitates simplifying assumptions like conditional independence and the computational time explodes when the dimension increases because the number of possible configurations becomes very high. The reader is referred to Müller and Czado (2017) for an example of the use of vine copulas in high dimensions.

This paper introduces the family of Fisher copulas in order to overcome the above-mentioned limitations of the available models; these limitations are especially problematic when the goal is to reproduce joint extreme precipitation events. As will be shown, this new copula family allows in particular for (i) high-dimensional modeling, (ii) non-vanishing upper tail dependence, (iii) radial asymmetry and (iv) a pairwise structure that helps to interpret results. The Fisher copula family can thus be seen as a very interesting alternative to the Normal, Student and chi-square classes of copulas, since none of these models possesses all these properties. Hence, the Fisher copulas are well adapted for the analysis of multi-site precipitation data, since such observations are typically tail asymmetric and show significant upper tail dependence. In addition, it could eventually be used in a spatial context since pairwise distances between stations can easily be incorporated into the model.

The remaining of the paper is structured as follows. In Section 2, the construction of the Fisher copula is detailed and theoretical properties are established. Two strategies for parameter estimation are proposed in Section 3 and their efficiency in moderate sample sizes is investigated in a simulation study. The usefulness of the newly introduced Fisher copulas is illustrated in Section 4 on precipitation data measured at 105 stations in the Aare river catchment in Switzerland; the results are compared to those obtained by Evin *et al.* (2017) using the Student copula. Technical details are relegated to an appendix.

2. CONSTRUCTION AND PROPERTIES OF THE FISHER COPULA

2.1. Some notes on the Student copula

As will be seen, the new Fisher copula that will be introduced is intimately related to the Student copula. Recall that a random vector $\mathbf{X} = (X_1, \ldots, X_d)$ is said to follow a d-variate Student distribution with correlation matrix Σ and $\nu \in \mathbb{N}$ degrees of freedom if it admits the stochastic representation

$$\mathbf{X} = \frac{1}{\sqrt{W/\nu}} \mathbf{Z},$$

where $\mathbf{Z}$ follows a d-variate Normal with correlation matrix Σ and W, which is independent of $\mathbf{Z}$, is distributed from the univariate chi-square law with ν degrees of freedom. In that case, the density of $\mathbf{X}$ is given for $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$ by

$$h^t_{\Sigma,\nu}(\mathbf{x}) = \frac{\Gamma\left(\frac{\nu+d}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\pi\nu)^{d/2}} |\Sigma|^{-1/2} \left(1 + \frac{\mathbf{x}\Sigma^{-1}\mathbf{x}^\top}{\nu}\right)^{-(\nu+d)/2},$$

where $\Gamma(s) = \int_0^\infty x^{s-1} e^{-x} \mathrm{d}x$ is the Gamma function. Generally speaking, the unique copula of a continuous random vector $\mathbf{X}$ can be extracted from the formula $C(u_1, \ldots, u_d) = \mathbb{P}\{X_1 \leq F_1^{-1}(u_1), \ldots, X_d \leq F_d^{-1}(u_d)\}$, where $(u_1, \ldots, u_d) \in [0,1]^d$. As a particular case, one deduces that the copula associated to the Student distribution can be implicitly expressed as

$$C^t_{\Sigma,\nu}(u_1, \ldots, u_d) = \int_{-\infty}^{F_\nu^{-1}(u_1)} \cdots \int_{-\infty}^{F_\nu^{-1}(u_d)} h^t_{\Sigma,\nu}(x_1, \ldots, x_d) \, \mathrm{d}x_d \cdots \mathrm{d}x_1,$$

where F_ν is the cumulative distribution function of the univariate Student distribution with ν degrees of freedom. Letting f_ν be the univariate Student density, the density of the Student copula can be expressed as

$$c^t_{\Sigma,\nu}(u_1,\ldots,u_d) = h^t_{\Sigma,\nu}\left\{F_\nu^{-1}(u_1),\ldots,F_\nu^{-1}(u_d)\right\} \Big/ \prod_{j=1}^{d} f_\nu\left\{F_\nu^{-1}(u_j)\right\}. \quad (2)$$

By construction, any marginal of dimension $p \in \{2,\ldots,d-1\}$ of the Student copula is a p-variate Student copula. Specifically, the copula of the subvector $\widetilde{\mathbf{X}} = (X_{i_1},\ldots,X_{i_p})$ is $C^t_{\widetilde{\Sigma},\nu}$, where $\widetilde{\Sigma}$ is the correlation matrix of $\widetilde{\mathbf{X}}$. In particular, the copula of the pair $(X_j, X_{j'})$ is the bivariate Student copula with correlation $\Sigma_{jj'}$; the latter will be noted $C^t_{\Sigma_{jj'},\nu}$ in the sequel.

As mentioned in the Introduction, the Student copula is radially symmetric. Formally, it means that if $(U_1,\ldots,U_d) \sim C^t_{\Sigma,\nu}$, then $(1-U_1,\ldots,1-U_d) \sim C^t_{\Sigma,\nu}$; details are given in Appendix A.1. A direct consequence is that $c^t_{\Sigma,\nu}(u_1,\ldots,u_d) = c^t_{\Sigma,\nu}(1-u_1,\ldots,1-u_d)$ for all $(u_1,\ldots,u_d) \in [0,1]^d$. Hence, the Student copula is tail symmetric in the sense that the lower and upper tails of its distribution are identical. This feature of the Student dependence structure can be restrictive in practice, especially in the case of precipitation data. Nevertheless, the Student copulas allow for non-null lower and upper tail dependence, since

$$\lambda_L\left(C^{\mathrm{t}}_{\Sigma_{jj'},\nu}\right) = \lambda_U\left(C^{\mathrm{t}}_{\Sigma_{jj'},\nu}\right) = 2\,F_{\nu+1}\left(-\sqrt{\frac{(\nu+1)(1-\Sigma_{jj'})}{1+\Sigma_{jj'}}}\right).$$

Finally note that from McNeil *et al.* (2005), Kendall's measure of association is such that

$$\tau\left(C^t_{\Sigma_{jj'},\nu}\right) = \frac{2}{\pi}\,\sin^{-1}\left(\Sigma_{jj'}\right),$$

2.2. From the Student to the Fisher copula

As initially suggested by Bàrdossy (2006), the chi-square copula family arises when one squares the components of a Normal random vector. Specifically, starting with a random vector $(Z_1,\ldots,Z_d)$ from the standard d-variate normal distribution with correlation matrix Σ, the chi-square copula with non-centrality parameter $a \geq 0$ and correlation matrix Σ is defined as the dependence structure of $((Z_1+a)^2,\ldots,(Z_d+a)^2)$. The properties of this family of copulas have been investigated by Quessy *et al.* (2016). In particular, one recovers the normal copula when $a \to \infty$, while for any $a \in [0,\infty)$, the chi-square copulas are radially asymmetric. However, the lower and upper tail dependence coefficients vanish.

Following a similar idea, a new copula family will be build upon considering Student random vectors instead of normal vectors. Specifically, let $\mathbf{X} = (X_1,\ldots,X_d)$ be a random vector distributed according to the Student distribution with correlation matrix Σ and $\nu \in \mathbb{N}$ degrees of freedom. Then, let $C^{\mathrm{F}}_{\Sigma,\nu}$ be the multivariate copula that corresponds to the dependence structure of $\mathbf{Y} = (X_1^2,\ldots,X_d^2)$. Because all the marginal distributions of $\mathbf{Y}$ are Fisher with 1 and ν degrees of freedom, this dependence model will be referred to the Fisher copula in the sequel. This family generalizes the class of centered chi-square dependence models since $C^{\mathrm{F}}_{\Sigma,\nu}$ tends to the centered chi-square copula as $\nu \to \infty$.

From similar arguments than those in Quessy *et al.* (2016), one can derive an expression for the Fisher copula and its associated density; this is given in Lemma 1.

Lemma 1 *For $\boldsymbol{\epsilon} = (\epsilon_1,\ldots,\epsilon_d)$, the d-dimensional Fisher copula can be expressed by*

$$C^{\mathrm{F}}_{\Sigma,\nu}(u_1,\ldots,u_d) = \sum_{\boldsymbol{\epsilon}\in\{-1,1\}^d}\left(\prod_{j=1}^{d}\epsilon_j\right) C^{\mathrm{t}}_{\Sigma,\nu}\left(\frac{1+\epsilon_1 u_1}{2},\ldots,\frac{1+\epsilon_d u_d}{2}\right),$$

where $C^{\mathrm{t}}_{\Sigma,\nu}$ is the Student copula. Its corresponding density can therefore be given by

$$c^{\mathrm{F}}_{\Sigma,\nu}(u_1,\ldots,u_d) = \frac{1}{2^d}\sum_{\boldsymbol{\epsilon}\in\{-1,1\}^d} c^{\mathrm{t}}_{\Sigma,\nu}\left(\frac{1+\epsilon_1 u_1}{2},\ldots,\frac{1+\epsilon_d u_d}{2}\right),$$

where $c^{\mathrm{t}}_{\Sigma,\nu}$ is the density of the Student copula.

An alternative expression for the Fisher copula density can be derived upon noting that $F_\nu^{-1}\{(1-u)/2\} = -F_\nu^{-1}\{(1+u)/2\}$. Hence, from Equation (2),

$$\begin{aligned}
&c^{\mathrm{t}}_{\Sigma,\nu}\left(\frac{1+\epsilon_1 u_1}{2},\ldots,\frac{1+\epsilon_d u_d}{2}\right)\\
&= h^t_{\Sigma,\nu}\left\{F_\nu^{-1}\left(\frac{1+\epsilon_1 u_1}{2}\right),\ldots,F_\nu^{-1}\left(\frac{1+\epsilon_d u_d}{2}\right)\right\}\Big/\prod_{j=1}^{d} f_\nu\left\{F_\nu^{-1}\left(\frac{1+\epsilon_1 u_j}{2}\right)\right\}\\
&= h^t_{\Sigma,\nu}\left\{\epsilon_1\,F_\nu^{-1}\left(\frac{1+u_1}{2}\right),\ldots,\epsilon_d\,F_\nu^{-1}\left(\frac{1+u_d}{2}\right)\right\}\Big/\prod_{j=1}^{d} f_\nu\left\{F_\nu^{-1}\left(\frac{1+u_j}{2}\right)\right\}.
\end{aligned}$$

From the conclusion of Lemma 1, the density of the Fisher copula can then be expressed as

$$c^{\mathrm{F}}_{\Sigma,\nu}(u_1,\ldots,u_d) = \left(\prod_{j=1}^{d} 2f_\nu\left\{F_\nu^{-1}\left(\frac{1+u_j}{2}\right)\right\}\right)^{-1} \times \sum_{\boldsymbol{\epsilon}\in\{-1,1\}^d} h^t_{\Sigma,\nu}\left\{\epsilon_1 F_\nu^{-1}\left(\frac{1+u_1}{2}\right),\ldots,\epsilon_d F_\nu^{-1}\left(\frac{1+u_d}{2}\right)\right\}. \quad (3)$$

This last formula will prove useful later in the context of parameter estimation.

2.3. Some properties of the Fisher copula

Using arguments similar as those given for the Student copula, one can conclude that the p-variate marginals of a Fisher copula also belong to the Fisher family. In particular, the copula of the pair (j, j') is the bivariate Fisher copula with parameter $\Sigma_{jj'}$; this is noted $C^{\mathrm{F}}_{\Sigma_{jj'},\nu}$ in the sequel.

Since the components of $(X_1,\ldots,X_d)$ can never be independent in the case of the Student distribution, even when $\Sigma = I$, the independence copula Π is not a member of the Fisher class of copulas. The only exception occurs when $\nu \to \infty$, in which case the Fisher copula tends to the centered chi-square and $\lim_{\nu\to\infty} C^{\mathrm{F}}_{I,\nu} = \Pi$. On the other side, letting Σ be such that $\Sigma_{jj'} = 1$ for all $j, j' \in \{1,\ldots,d\}$ corresponds to the situation where $X_1 = \cdots = X_d$ almost surely, and then also $X_1^2 = \cdots = X_d^2$ almost surely. As a consequence, $C^{\mathrm{F}}_{\Sigma,\nu} = M$ is the Fréchet–Hoeffding upper bound associated to the d-variate perfect positive dependence.

It does not seem possible to derive a closed-form expression for Kendall's tau associated to the Fisher copula. Nevertheless, the latter can be approximated by a Monte–Carlo simulation based on the empirical version of Kendall's tau, which from a sample of pairs $(X_{11}, X_{12}),\ldots,(X_{n1}, X_{n2})$ is defined by

$$\widehat{\tau} = -1 + \frac{4}{n(n-1)} \sum_{1\le i<i'\le n} \mathbb{I}\{(X_{i1} - X_{i'1})(X_{i2} - X_{i'2}) > 0\}. \quad (4)$$

Curves of $\tau(C^{\mathrm{F}}_{\rho,\nu})$ as a function of ρ, as estimated with the help of 1,000 samples of size $n = 500$ from the Fisher copula, are given in Figure 1 for $\nu \in \{1, 2, 3, 4, 10\}$. Note that as ν increases, $\tau(C^{\mathrm{F}}_{\rho,\nu})$ will tend to the simple expression derived by Quessy *et al.* (2016) in the case of the (centered) chi-square copula, namely $\lim_{\nu\to\infty} \tau(C^{\mathrm{F}}_{\rho,\nu}) = \{(2/\pi)\sin^{-1}\rho\}^2$. Otherwise, somewhat as a consequence of the fact that $\rho = 0$ does not correspond to independence, Kendall's tau at the origin does not vanish.

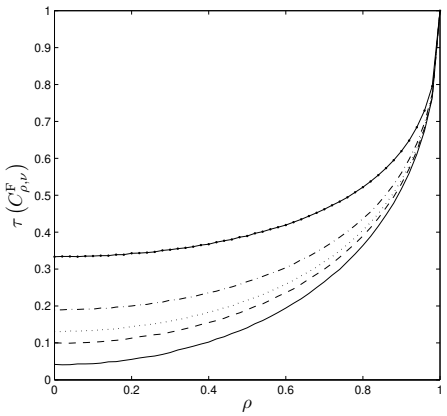


Figure 1. Curves of Kendall's tau as a function of $\rho \in [0, 1]$ for the Fisher copula $C^{\mathrm{F}}_{\rho,\nu}$ with $\nu = 1$ (solid line with dots), $\nu = 2$ (dotted line), $\nu = 3$ (dots), $\nu = 4$ (dashed line) and $\nu = 10$ (solid line)

One key feature of the members of the Fisher family is that unlike those in the Student family, they are radially asymmetric. This characteristic can be noticed in Figure 2, where one can find pair plots from 1,000 independent copies of a triplet (X_1, X_2, X_3) whose dependence structure is the Fisher copula with $\nu = 3$ degrees of freedom and where Kendall's tau of the pairs have been set to $\tau(X_1, X_2) = 0.25$, $\tau(X_1, X_3) = 0.50$ and $\tau(X_2, X_3) = 0.75$; these values correspond approximately to $\Sigma_{12} = 0.58$, $\Sigma_{13} = 0.88$ and $\Sigma_{23} = 0.98$. A formal statement of this property is given later in Proposition 2.

The following result, whose proof is given in Appendix A.3, provides expressions for the lower and upper tail dependence coefficients of the Fisher copula.

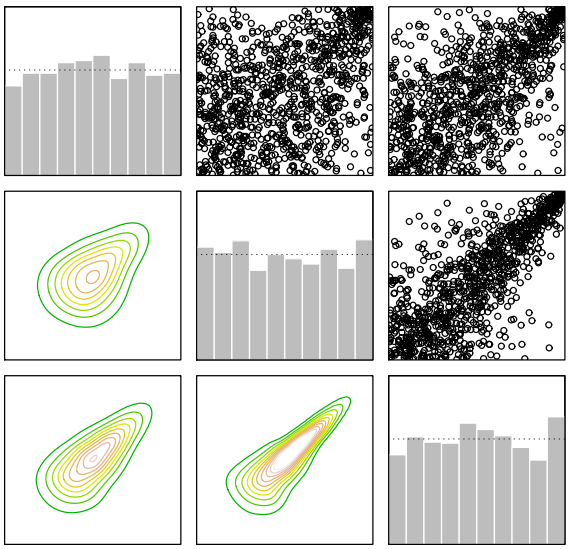


Figure 2. Pair plots of 1,000 independent copies of a triplet (X_1, X_2, X_3) whose dependence structure is the Fisher copula with $\nu = 3$ degrees of freedom and $\tau(X_1, X_2) = 0.25$, $\tau(X_1, X_3) = 0.50$ and $\tau(X_2, X_3) = 0.75$, and whose marginal distributions is uniform on $[0, 1]$ (upper triangle) and the normalized contour plots of these data (lower triangle)

Proposition 1 *The tail dependence coefficients of the Fisher copula $C^{\mathrm{F}}_{\Sigma,\nu}$ are given for each $j < j' \in \{1, \ldots, d\}$ by $\lambda_L(C^{\mathrm{F}}_{\Sigma_{jj'},\nu}) = 0$ and*

$$\begin{aligned}\lambda_U\left(C^{\mathrm{F}}_{\Sigma_{jj'},\nu}\right) &= \lambda_U\left(C^{\mathrm{t}}_{\Sigma_{jj'},\nu}\right) + \lambda_U\left(C^{\mathrm{t}}_{-\Sigma_{jj'},\nu}\right) \\ &= 2\left\{F_{\nu+1}\left(-\sqrt{\frac{(\nu+1)(1-\Sigma_{jj'})}{1+\Sigma_{jj'}}}\right) + F_{\nu+1}\left(-\sqrt{\frac{(\nu+1)(1+\Sigma_{jj'})}{1-\Sigma_{jj'}}}\right)\right\}.\end{aligned}$$

In particular, $\lim_{\nu\to\infty} \lambda_L(C^{\mathrm{F}}_{\Sigma_{jj'},\nu}) = \lim_{\nu\to\infty} \lambda_U(C^{\mathrm{F}}_{\Sigma_{jj'},\nu}) = 0$, recovering a result obtained by Quessy et al. (2016) for the centered chi-square copula.

Proposition 1 ensures that the Fisher copulas allow for upper tail dependence; this is an interesting characteristic when compared, for instance, to the chi-square copulas. Values of $\lambda_U(C^{\mathrm{F}}_{\rho,\nu})$ are provided in Table 1 for some selections of $\rho \in [0, 1]$ and $\nu \in \mathbb{N}$. Note that since the number of degrees of freedom is the same for all the pairs in the multivariate Fisher copula, a consequence of Proposition 1 is that the level of upper tail dependence for a given pair is managed by the corresponding entry of the correlation matrix.
Another consequence of Proposition 1 is that unlike the Normal, Student and other elliptical copulas, the d-dimensional Fisher copula is radially asymmetric; this is formally stated in Proposition 2.

Proposition 2 *For any positive definite matrix $\Sigma \in \mathbb{R}^{d\times d}$ and $\nu < \infty$, the d-variate Fisher copula $C^{\mathrm{F}}_{\Sigma,\nu}$ is radially asymmetric.*

3. PARAMETER ESTIMATION

Consider n independent copies $\mathbf{X}_1, \ldots, \mathbf{X}_n$ of a random vector $\mathbf{X} = (X_1, \ldots, X_d)$ whose distribution is assumed to have continuous marginal distributions $F_1, \ldots, F_d$ and a copula that belongs to the family of d-variate Fisher copulas $\{C^{\mathrm{F}}_{\Sigma,\nu}; \Sigma, \nu\}$. The goal of this section is to develop efficient strategies for the estimation of the unknown parameters Σ and ν in a context where the marginal distributions are unknown; to this end, semi-parametric inference methods will be adopted in the sequel.

3.1. The case when ν is fixed: inversion of Kendall's tau

As noted in Section 2.3, each entry of the correlation matrix Σ can be related implicitly to the value of Kendall's tau of the corresponding pair. In other words, there is a function $g_\nu : [0, 1] \to [0, 1]$ such that $\Sigma_{jj'} = g_\nu\{\tau(X_j, X_{j'})\}$ for each $j \neq j'$. If the value of $\nu \in \mathbb{N}$ is assumed to be known, one can then estimate the entries of Σ with $\widehat{\Sigma}_{jj'}(\nu) = g_\nu(\widehat{\tau}_{jj'})$, where $\widehat{\tau}_{jj'}$ is

Table 1. Some values of the upper tail dependence coefficient of the bivariate Fisher copula $C^{\mathrm{F}}_{\rho,\nu}$

	Degrees of freedom							
ρ	$\nu=1$	$\nu=2$	$\nu=3$	$\nu=4$	$\nu=5$	$\nu=10$	$\nu=15$	$\nu=30$
0.00	0.586	0.363	0.232	0.151	0.100	0.014	0.002	0.000
0.10	0.588	0.367	0.236	0.155	0.104	0.016	0.003	0.000
0.20	0.593	0.376	0.248	0.168	0.116	0.022	0.005	0.000
0.30	0.602	0.393	0.269	0.190	0.138	0.034	0.010	0.000
0.40	0.616	0.417	0.298	0.222	0.170	0.053	0.019	0.001
0.50	0.634	0.449	0.338	0.265	0.212	0.082	0.035	0.003
0.60	0.658	0.491	0.390	0.321	0.269	0.126	0.063	0.009
0.70	0.691	0.545	0.457	0.394	0.344	0.191	0.112	0.026
0.80	0.735	0.618	0.545	0.491	0.446	0.293	0.201	0.073
0.90	0.802	0.723	0.671	0.630	0.595	0.463	0.372	0.211
0.92	0.820	0.750	0.705	0.667	0.635	0.512	0.426	0.264
0.94	0.842	0.783	0.743	0.710	0.682	0.571	0.492	0.335
0.96	0.869	0.822	0.789	0.762	0.738	0.645	0.576	0.432
0.98	0.905	0.873	0.851	0.831	0.814	0.745	0.693	0.580
1.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

the empirical Kendall's tau defined in (4) computed from the sample of pairs $(X_{1j}, X_{1j'}), \ldots, (X_{nj}, X_{nj'})$. Since g_ν has no explicit expression, the latter is approximated from 1,000 samples of size $n = 500$, as was the case for the curves in Figure 1.

The results of an investigation on the performance of this estimator in the light of estimated relative root mean-squared error and relative bias are to be found in Table 2. Recall that for an estimator $\widehat{\theta}$ of a real-valued parameter $\theta > 0$, the relative root mean-squared error and relative bias are defined respectively by

$$\mathrm{RRMSE}(\widehat{\theta}) = \sqrt{\mathrm{E}\left\{\left(\frac{\widehat{\theta}}{\theta} - 1\right)^2\right\}} \quad \text{and} \quad \mathrm{RB}(\widehat{\theta}) = \mathrm{E}\left(\frac{\widehat{\theta}}{\theta} - 1\right).$$

In order to ease the interpretation of the results, the model has been parametrized in terms of $\tau(C^{\mathrm{F}}_{\rho,\nu}) \in \{.25, .50, .75\}$ corresponding to low, moderate and high levels of dependence. Looking at Table 2, one observes that the accuracy of $\widehat{\rho}$ increases as the sample size increases, as expected. Generally speaking, $\mathrm{RB}(\widehat{\rho})$ and $\mathrm{RRMSE}(\widehat{\rho})$ decrease as $\tau(C^{\mathrm{F}}_{\rho,\nu})$ (or similarly ρ) increases. The estimator tends to be better as the number of degrees of freedom increases, but the results are nevertheless quite similar for $\nu \in \{6, 9, 12, 15, 18\}$. Overall, the inversion of Kendall's tau estimator $\widehat{\rho}$ performs well.

3.2. The pseudo maximum likelihood estimator and its limitations

In many applications, the number of degrees of freedom $\nu \in \mathbb{N}$ is unknown. In that case, an alternative to the inversion of Kendall's tau is the pseudo maximum likelihood estimator proposed and investigated by Genest *et al.* (1995). The latter is a natural rank-based version of the maximum likelihood estimator in which the unknown univariate distributions $F_1, \ldots, F_d$ are estimated by their natural nonparametric counterparts provided by the (re-scaled) empirical distribution functions $F_{n1}, \ldots, F_{nd}$, where for each $j \in \{1, \ldots, d\}$,

$$F_{nj}(x) = \frac{1}{n+1} \sum_{i=1}^{n} \mathbb{I}(X_{ij} \leq x).$$

Note that dividing by $n+1$ instead of n avoids potential numerical problems when evaluating copula densities when one or more components is 1. In the case of the Fisher copula, letting $\mathcal{A}_{20} = \{1, \ldots, 20\}$ be the set of the first twenty integers and $\mathcal{M}_d$ the set of d-dimensional correlation matrices, the pseudo maximum likelihood (PML) estimator is defined by

$$(\widehat{\Sigma}, \widehat{\nu}) = \underset{(\Sigma,\nu)\in\mathcal{M}_d\times\mathcal{A}_{20}}{\operatorname{argmax}} \sum_{i=1}^{n} \ln c^{\mathrm{F}}_{\Sigma,\nu}\left\{F_{n1}(X_{i1}), \ldots, F_{nd}(X_{id})\right\}. \tag{5}$$

From Equation (3) and defining, for each $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, d\}$,

$$\widehat{X}_{ij}(\nu) = F_\nu^{-1}\left\{\frac{1 + F_{nj}(X_{ij})}{2}\right\},$$

Table 2. Estimation, based on 1,000 replicates, of the relative root mean-squared error (RRMSE) and relative bias (RB) (in %) of the inversion of the Kendall's tau estimator of ρ for the bivariate Fisher copula $C^{\mathrm{F}}_{\rho,\nu}$

			$n=100$		$n=200$		$n=400$	
ν	$\tau(C^{\mathrm{F}}_{\rho,\nu})$	ρ	RB($\widehat{\rho}$)	RRMSE($\widehat{\rho}$)	RB($\widehat{\rho}$)	RRMSE($\widehat{\rho}$)	RB($\widehat{\rho}$)	RRMSE($\widehat{\rho}$)
	0.25	.582	-8.99	35.86	-5.11	24.50	-3.64	15.27
3	0.50	.876	-0.69	4.85	-0.36	3.28	-0.32	2.24
	0.75	.976	-0.00	0.71	0.02	0.49	0.01	0.38
	0.25	.659	-2.83	17.91	-1.60	12.15	-0.78	7.45
6	0.50	.888	-0.40	3.72	-0.20	2.46	-0.14	1.71
	0.75	.977	-0.04	0.57	-0.02	0.40	-0.01	0.24
	0.25	.679	-3.21	16.91	-1.80	10.52	-0.90	7.20
9	0.50	.891	-0.09	3.21	-0.18	2.26	-0.05	1.57
	0.75	.977	-0.12	0.62	-0.07	0.39	-0.09	0.28
	0.25	.685	-1.92	14.62	-0.87	9.10	-0.83	6.99
12	0.50	.892	-0.44	3.27	-0.26	2.23	-0.13	1.61
	0.75	.977	-0.05	0.60	-0.02	0.38	-0.02	0.29
	0.25	.691	-2.36	14.44	-1.34	9.07	-0.62	5.67
15	0.50	.893	-0.29	3.12	-0.21	2.20	-0.13	1.43
	0.75	.977	0.02	0.55	0.08	0.37	0.09	0.26
	0.25	.694	-1.68	12.97	-1.29	8.70	-0.42	5.67
18	0.50	.894	-0.07	2.90	-0.12	2.01	0.02	1.39
	0.75	.977	-0.13	0.65	-0.08	0.42	-0.09	0.31

one can write alternatively that

$$\begin{aligned}(\widehat{\Sigma},\widehat{\nu}) &= \underset{(\Sigma,\nu)\in\mathcal{M}_d\times\mathcal{A}_{20}}{\operatorname{argmax}} \sum_{i=1}^{n} \ln\left\{\left(\prod_{j=1}^{d} 2f_\nu(\widehat{X}_{ij}(\nu))\right)^{-1} \sum_{\boldsymbol{\epsilon}\in\{-1,1\}^d} h^t_{\Sigma,\nu}\left(\epsilon_1\widehat{X}_{i1}(\nu),\ldots,\epsilon_d\widehat{X}_{id}(\nu)\right)\right\} \\ &= \underset{(\Sigma,\nu)\in\mathcal{M}_d\times\mathcal{A}_{20}}{\operatorname{argmax}} \sum_{i=1}^{n}\left[\ln\left\{\sum_{\boldsymbol{\epsilon}\in\{-1,1\}^d} h^t_{\Sigma,\nu}\left(\epsilon_1\widehat{X}_{i1}(\nu),\ldots,\epsilon_d\widehat{X}_{id}(\nu)\right)\right\} - \sum_{j=1}^{d}\ln f_\nu(\widehat{X}_{ij}(\nu))\right].\end{aligned}$$

Computing $(\widehat{\Sigma},\widehat{\nu})$ above is however quite unrealistic in practice, especially when d is large, since the expression inside the brackets of the last equation involves 2^d summands. Moreover, the maximization over the product-set $\mathcal{M}_d\times\mathcal{A}_{20}$ is very time-consuming. For these reasons, more computationally efficient strategies based on a two-step approach will be considered.

3.3. Two-step pseudo maximum likelihood estimators

The computational efficiency of $(\widehat{\Sigma},\widehat{\nu})$ defined in (5) can be significantly improved by using a two-step procedure based on the relationship between each entry of Σ and the pairwise values of Kendall's tau. Specifically, one can first condition on $\Sigma^\star(\nu)\in\mathbb{R}^{d\times d}$, with $\Sigma^\star_{jj'}(\nu)=g_\nu(\widehat{\tau}_{jj'})$, an estimate ν with

$$\widehat{\nu}_1 = \underset{\nu\in\mathcal{A}_{20}}{\operatorname{argmax}} \sum_{i=1}^{n}\left[\ln\left\{\sum_{\boldsymbol{\epsilon}\in\{-1,1\}^d} h^t_{\Sigma^\star(\nu),\nu}\left(\epsilon_1\widehat{X}_{i1}(\nu),\ldots,\epsilon_d\widehat{X}_{id}(\nu)\right)\right\} - \sum_{j=1}^{d}\ln f_\nu(\widehat{X}_{ij}(\nu))\right]. \tag{6}$$

Then, each entry $\Sigma_{jj'}$ of Σ is estimated with $(\widehat{\Sigma}_1)_{jj'}=g_{\widehat{\nu}_1}(\widehat{\tau}_{jj'})$. Note that similarly as for the inversion of Kendall's tau estimator, g_ν is replaced by an estimated version.

The computation of $\widehat{\nu}_1$ is based on the full d-dimensional density of the Fisher copula. Similarly as the PML estimator, the latter involves summing 2^d multivariate densities and thus can become very heavy to compute as d increases. In order to overcome this issue, an alternative approach consists in basing the estimator on the pairwise likelihood; see Varin *et al.* (2011) for more details on this kind of estimators, and more generally on composite likelihoods. Specifically, instead of using the full d-variate copula density, one considers

$$\begin{aligned}\widehat{\nu}_2 = \underset{\nu\in\mathcal{A}_{20}}{\operatorname{argmax}} \sum_{i=1}^{n}\sum_{1\le j<j'\le d}\Bigg[&\ln\left\{\sum_{(\epsilon_1,\epsilon_2)\in\{-1,1\}^2} h^t_{g_\nu(\widehat{\tau}_{jj'}),\nu}\left(\epsilon_1\widehat{X}_{ij}(\nu),\epsilon_2\widehat{X}_{ij'}(\nu)\right)\right\} \\ &- \ln f_\nu(\widehat{X}_{ij}(\nu)) - \ln f_\nu(\widehat{X}_{ij'}(\nu))\Bigg],\end{aligned} \tag{7}$$

where $h^t_{\rho,\nu}$ is the bivariate Student density with ν degrees of freedom and correlation ρ. Similarly as for $\widehat{\nu}_1$, each entry $\Sigma_{jj'}$ of Σ is estimated with $(\widehat{\Sigma}_2)_{jj'} = g_{\widehat{\nu}_2}(\widehat{\tau}_{jj'})$. While the computation of $\widehat{\nu}_2$ involves 2^d summands, computing $\widehat{\nu}_1$ only requires the summation of $d(d-1)/2$ bivariate Fisher densities. This provides a very significant computational advantage, especially when d becomes large; in fact, $\widehat{\nu}_1$ is just impracticable when $d \geq 10$.

A first numerical investigation on the performance of these estimators has been made for the bivariate Fisher copula $C^{\mathrm{F}}_{\rho,\nu}$; in that case, the two proposed strategies yield the same estimators, *i.e.* $\widehat{\rho}_1 = \widehat{\rho}_2 = \widehat{\rho}$ and $\widehat{\nu}_1 = \widehat{\nu}_2 = \widehat{\nu}$. The results reported in Table 3 concern RRMSE($\widehat{\rho}$) and RB($\widehat{\rho}$), as well as RRMSE($\widehat{\nu}$) and RB($\widehat{\nu}$), as estimated from 1,000 samples from the Fisher copula. Here again, the latter has been parametrized in terms of $\tau(C^{\mathrm{F}}_{\rho,\nu}) \in \{.25, .50, .75\}$ corresponding to low, moderate and high levels of dependence. As could have been expected, the values of the relative root mean-squared error (RRMSE) and relative bias (RB) for both $\widehat{\rho}$ and $\widehat{\nu}$ tend to decrease as the sample size increases. Generally speaking, the RRMSE and RB decrease significantly as the value of Kendall's tau increases. Note also that the accuracy of $\widehat{\nu}$ and $\widehat{\rho}$ is better when $\nu = 10$ compared to $\nu = 5$, both in terms of RRMSE and RB, except for $\widehat{\rho}$ when $\tau(C^{\mathrm{F}}_{\rho,\nu}) = 0.25$. Globally, the estimation of ν is comparatively less accurate than the estimation of ρ.

Table 3. Estimation, based on 1,000 replicates, of the relative root mean-squared error (RRMSE) and relative bias (RB) (in %) of the two-step pseudo-maximum likelihood estimators of ρ and ν for the bivariate Fisher copula $C^{\mathrm{F}}_{\rho,\nu}$

				Estimation of ρ		Estimation of ν	
Sample size	ν	$\tau(C^{\mathrm{F}}_{\rho,\nu})$	ρ	RRMSE($\widehat{\rho}$)	RB($\widehat{\rho}$)	RRMSE($\widehat{\nu}$)	RB($\widehat{\nu}$)
250	5	.25	.645	35.88	-17.07	133.13	38.08
		.50	.886	7.25	-2.29	110.44	30.24
		.75	.977	0.46	-0.06	94.51	32.76
	10	.25	.681	37.63	-19.69	72.67	-8.64
		.50	.892	5.10	-1.20	65.50	7.37
		.75	.977	0.38	-0.05	59.05	11.94
500	5	.25	.645	29.25	-14.22	106.58	19.64
		.50	.886	4.01	-0.86	76.38	17.52
		.75	.977	0.29	-0.03	68.67	21.58
	10	.25	.681	28.63	-13.59	67.27	-4.35
		.50	.892	3.04	-0.51	55.89	7.89
		.75	.977	0.26	-0.03	53.94	15.36
1,000	5	.25	.645	21.88	-9.89	91.54	17.12
		.50	.886	2.08	-0.29	48.65	12.94
		.75	.977	0.21	-0.02	36.48	9.02
	10	.25	.681	19.89	-7.71	61.47	3.71
		.50	.892	1.05	-0.11	47.97	10.68
		.75	.977	0.17	-0.01	44.63	11.93

A second numerical investigation concerns the two-step estimators $\widehat{\nu}_1$ and $\widehat{\nu}_2$ of the degree of freedom in the d-dimensional Fisher copula defined in Equation (6) and Equation (7), respectively; the results can be found in Table 4 when $d \in \{3, 4, 5, 10\}$. The entries of the correlation matrix Σ have been chosen in such a way that the $K = d(d-1)/2$ values of the pairwise Kendall's tau cover the set $\{(k-1)/K; k = 1, \ldots, K\}$. Specifically, for $g_\nu : [0,1] \to [0,1]$ such that $\Sigma_{jj'} = g_\nu\{\tau(X_j, X_{j'})\}$,

$$\Sigma_{12} = g_\nu(0) = 0, \quad \Sigma_{13} = g_\nu\left(\frac{1}{K}\right), \quad \ldots, \quad \Sigma_{d-1,d} = g_\nu\left(\frac{K-1}{K}\right).$$

This construction allows to cover a large spectrum of levels of pairwise dependence. For example, when $d = 3$, one has $\Sigma_{12} = 0$, $\Sigma_{13} = g_\nu(1/3)$ and $\Sigma_{23} = g_\nu(2/3)$. When the resulting matrix Σ is not positive definite, as required, a slightly modified version $\widetilde{\Sigma}$ is used instead, following the idea of Higham (2002) that consists in replacing negative eigenvalues in the diagonal matrix by some small value. Looking at Table 4, one observes that as expected, RRMSE($\widehat{\nu}_1$), RB($\widehat{\nu}_1$), RRMSE($\widehat{\nu}_2$) and RB($\widehat{\nu}_2$) decrease as the sample size n increases. Also, similarly as in the bivariate case (see results in Table 3), the estimation of ν

is more accurate when $\nu = 10$ compared to $\nu = 5$, particularly in terms of relative biases; this behavior is more noticeable for $\widehat{\nu}_2$. In addition, note that RRMSE($\widehat{\nu}_1$) and RB($\widehat{\nu}_1$) increase as the dimension d increases. The performance of $\widehat{\nu}_2$ as a function of d is less clear; indeed, its behavior could probably be explained by the fact that the loss of information due to the use of the pairs instead of the whole d-dimensional distribution is somewhat counterbalanced by the increase of information provided by the dimension. Finally observe that the relative biases of $\widehat{\nu}_2$ are high compared to those of $\widehat{\nu}_1$.

Globally, the accuracy of $\widehat{\nu}_2$ in terms of RRMSE is rather comparable to that of $\widehat{\nu}_1$, except maybe when $\nu = 5$ and $d = 3$. Considering that the former is more computationally efficient, especially as d becomes large, the use of $\widehat{\nu}_2$ can be recommended as long as the dimension d is relatively high, *e.g.* $d \geq 5$. Indeed, the use of $\widehat{\nu}_1$ becomes unrealistic when $d \geq 10$; this is the reason why in Table 4, the results for $\widehat{\nu}_1$ when $d = 10$ are not available.

Table 4. Estimation, based on 1,000 replicates, of the relative root mean-squared error (RRMSE) and relative bias (RB) (in %) of the two-step pseudo-maximum likelihood estimators $\widehat{\nu}_1$ and $\widehat{\nu}_2$ of the degree of freedom for the multivariate Fisher copula

			RRMSE		RB	
Sample size	ν	d	$\widehat{\nu}_1$	$\widehat{\nu}_2$	$\widehat{\nu}_1$	$\widehat{\nu}_2$
250	5	3	67.80	100.68	11.90	37.82
		4	73.21	78.38	19.16	28.48
		5	76.07	73.67	23.42	28.10
		10	—	80.34	—	31.86
	10	3	50.72	57.14	-2.60	12.81
		4	50.74	49.91	1.44	6.42
		5	52.75	53.74	-1.19	12.43
		10	—	52.34	—	5.48
500	5	3	44.66	82.30	5.94	30.00
		4	53.29	52.72	9.10	17.60
		5	56.24	54.41	16.56	21.04
		10	—	73.15	—	36.22
	10	3	44.41	49.29	0.66	11.98
		4	44.94	43.72	0.49	6.04
		5	47.11	47.29	3.13	15.84
		10	—	51.78	—	19.76

4. ILLUSTRATION ON MULTI-SITE PRECIPITATION DATA

The usefulness of the multivariate Fisher copula will be illustrated here on precipitation measurements made at $d = 105$ stations located within or close to the Aare river catchment in Switzerland. These stations cover relatively well the Aare river catchment, as can be seen from Figure 1 in Evin *et al.* (2017). More precisely, stations at daily time step having less than three years of missing data during the period 1930–2014 have been included in the database and the 3-day precipitation amounts for the month of September from the years 1930 to 2014 have been considered, for a total of $n = 849$ observations. It is worth noting that there is no clear pattern of spatial dependence in these data, as understood in the classical sense where the strength of the dependence between two stations is a monotone decreasing function of their distance. This is therefore a case of a pure modeling of high-dimensional data.

These data will first be filtered following the approach of Evin *et al.* (2017). Specifically, define the vector $\mathbf{Y}_t = (Y_t(1), \ldots, Y_t(d))$ of the 3-day precipitation amount for the d stations at time $t \in \mathbb{Z}$. The process $(\mathbf{Y}_t)_{t\in\mathbb{Z}}$ is then modeled with the help of an auxiliary vector autoregressive model of order 1 given by $\mathbf{Z}_t = A\,\mathbf{Z}_{t-1} + \boldsymbol{\epsilon}_t$, where $\mathbf{Z}_t = (Z_t(1), \ldots, Z_t(d))$ is standard multivariate Normal, $\boldsymbol{\epsilon}_t$ is a d-dimensional vector of zero-mean innovations that is independent of $\mathbf{Z}_{t-1}$, and $A \in \mathbb{R}^{d\times d}$ is a diagonal matrix whose elements are the lag-1 serial correlation coefficients. The components of $\mathbf{Y}_t$ are then defined by $Y_t(k) = F_k^{-1} \circ \Phi\{Z_t(k)\}$, so that for each $t \in \mathbb{Z}$, $\mathbf{Y}_t$ has marginal distributions $F_1, \ldots, F_d$ and a copula that belongs to the Normal family. Finally, consider for each $t \in \mathbb{Z}$ the residual $\mathbf{e}_t = \widehat{\mathbf{Z}}_t - \widehat{A}\,\widehat{\mathbf{Z}}_{t-1}$, where $\widehat{A}$ is an estimation of the unknown matrix A and the components of $\widehat{\mathbf{Z}}_t$ are $\widehat{Z}_t(k) = \Phi^{-1} \circ F_{nk}\{Y_t(k)\}$, with F_{nk} being the marginal empirical distribution of component $k \in \{1, \ldots, d\}$.

The goal of Evin *et al.* (2017) was to model the copula of the process of innovations $(\boldsymbol{\epsilon}_t)_{t\in\mathbb{Z}}$ from the sample $e_2, \ldots, e_n$ of residuals. To this end, they assumed a Student copula for the dependence structure and obtained $\widehat{\nu} = 20$ from the pseudo-likelihood method. However, as can be seen from the plots of standardized ranks in Figure 3 for three pairs of stations that are representative of the whole data set, these residuals show clear evidence of tail asymmetry, which discards the Student copula as an appropriate model for these data. In addition, the boxplot of the pairwise nonparametric estimators of the lower tail dependence coefficient in Figure 4, using the estimator of Schmidt and Stadtmüller (2006), suggests that most of the pairs have a lower tail dependence coefficient that is close to zero. Note that this conclusion took into account the significant bias of this estimator when there is no tail dependence, as well documented by Serinaldi *et al.* (2015). On the other side, the boxplot of the upper tail dependence coefficients suggests that most of the pairs have significant positive upper tail dependence. These features of the data set are arguments in favour of the use of the Fisher family instead of the Student or chi-square classes of copulas, for example. Details on the tail dependence coefficients estimators of Schmidt and Stadtmüller (2006) are to be found in Appendix A.5.

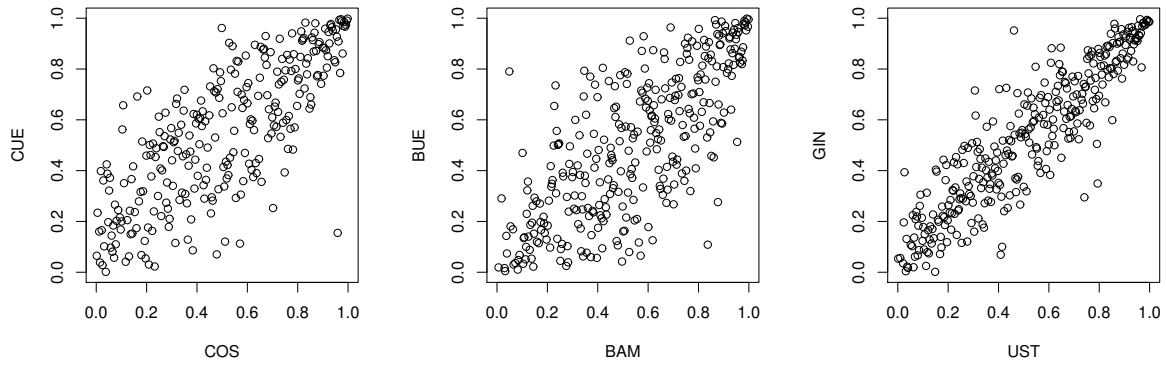


Figure 3. Scatter plots of the standardized ranks for three selected pairs of stations for the Aare river precipitation data

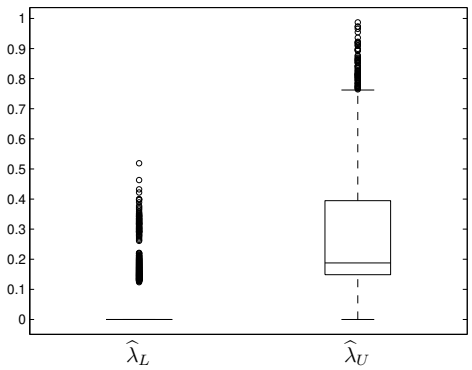


Figure 4. Boxplots of the nonparametric estimations of the lower and upper tail dependence coefficients for the 5 460 pairs in the Aare river precipitation data

Under an assumed multivariate Fisher copula, the two-step pairwise estimator yields $\widehat{\nu}_2 = 15$ for the degree of freedom. Note that the computation of $\widehat{\nu}_1$ would have been totally unrealistic here, since the likelihood function would involve summing $2^{105} \approx 4 \times 10^{31}$ multivariate Student densities; comparatively, the computation of $\widehat{\nu}_2$ required summing $105(104)/2 = 5,460$ bivariate Fisher densities. It can be seen in Figure 5 that fitting the Fisher copula succeeds in reproducing the tail behavior of the residuals when compared to the fit provided by the Student copula. Indeed, these figures show the boxplots of the difference between the nonparametric estimator of the tail dependence coefficients and its semi-parametric version (*i.e.* with estimated parameters) under an assumed Student or Fisher copula. It seems clear that the Fisher copula is much more accurate at predicting the values of the upper tail dependence coefficient for the pairs compared to the Student copula; indeed, the mean difference is 0.013 for the Fisher copula and 0.210 for the Student copula. For the lower tail dependence, the mean difference is 0.042 for the Fisher copula and -0.008 for the Student copula.

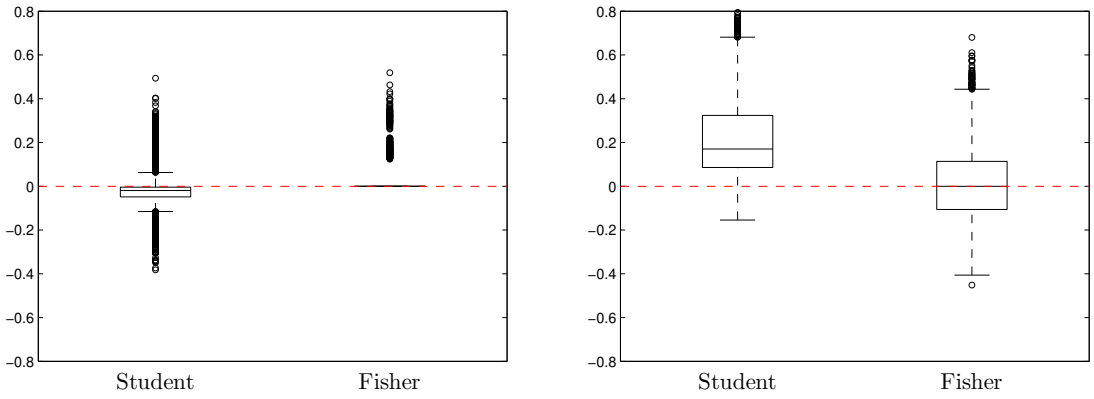


Figure 5. Boxplots of the difference between fully nonparametric estimations of the tail dependence coefficients and semi-parametric estimations under an assumed multivariate Student or Fisher copula for the 5 460 pairs in the Aare river precipitation data; left panel: lower tail dependence; right panel: upper tail dependence

5. CONCLUSION AND FUTURE INVESTIGATIONS

A new family of multidimensional dependence models, called the Fisher copulas, has been introduced in this work. This class of dependence structures arises as the set of copulas that can be extracted from the distribution of Student random vectors whose components are squared; the idea is similar to Bàrdossy (2006) for the construction of the chi-square copula family. Indeed, the Fisher copula family generalizes the class of centered chi-square copulas. One of the main features of the Fisher copulas is their radial asymmetry, *i.e.* unlike the Student copulas, their lower and upper tails are not identical. In addition, explicit expressions have been derived for the tail dependence coefficients, showing that the Fisher copulas have no lower tail dependence, but positive upper tail dependence. These key characteristics of the Fisher copula family are very interesting from a practical point-of-view.

A crucial point for the applicability of the Fisher copulas is to be able to efficiently estimate their parameters; this task is rather challenging, especially when the dimension d is large. Indeed, the popular and natural pseudo maximum likelihood (PML) estimator is computationally inefficient for the Fisher copulas, since its computation requires an optimization over the whole set of correlation matrices, while involving 2^d terms for the computation of the density. For these reasons, two-step procedures have been proposed as more computationally efficient alternatives to the PML estimator. The idea is to exploit a relationship that exists between the pairwise Kendall's tau and the entries of the correlation matrix, so that the latter is first estimated from the empirical Kendall's tau of all pairs and then, conditionally on this estimation, the degree of freedom is estimated from a simple optimization on $\mathbb{N}$. Two variants have been considered, namely one using the whole copula density and another based on a pairwise likelihood. With the help of simulations, it has been established that the latter is far more computationally efficient when d becomes large, while being generally comparable to the former in lower dimensions.

The usefulness of the newly introduced Fisher copula family has then been illustrated on the modeling of precipitation data observed at $d = 105$ stations within or close to the Aare river catchment in Switzerland. The goal was to capture the observed asymmetries in the tails and the significant upper tail dependence featured by the residuals extracted from a semi-parametric linear model considered by Evin *et al.* (2017). While the Student copula family somewhat fails at capturing these important features, the newly introduced Fisher family proves especially suitable in that case. It is thus expected that this modeling taking into consideration tail asymmetry and positive upper-tail dependence will lead to a better reproduction of extremes using the stochastic precipitation model of Evin *et al.* (2017). This expected improvement over the use of the Student copula family could be more formally established using the comprehensive and systematic evaluation framework of Bennett (2017).

In future investigations, it could be interesting to add even more flexibility to the Fisher copula family. Specifically, letting $\mathbf{X} = (X_1, \ldots, X_d)$ be a random vector from the Student distribution with $\nu \in \mathbb{N}$ degrees of freedom and correlation matrix Σ, one could consider the copula $C^{\mathrm{F}}_{\Sigma,\nu,a}$ of $\mathbf{Y} = ((X_1 + a)^2, \ldots, (X_d + a)^2)$, where $a \geq 0$ is a non-centrality parameter. This construction would create a *continuum* between the Fisher and Student families, since on one side, $a = 0$ corresponds to the Fisher copulas, and on the other side, letting $a \to \infty$ allows to recover the Student dependence structures. However, this gain in flexibility would raise at the same time the issue of estimating a, which is far from being straightforward. Finally note that other multivariate copula models could be build by squaring the components of elliptical random vectors other than the Student, *e.g.* Laplace or logistic.

APPENDIX

A. TECHNICAL DETAILS

A.1. Radial symmetry of the Student copula

For a random vector $\mathbf{X}$ distributed from the Student distribution, let $(U_1, \ldots, U_d) = (F_\nu(X_1), \ldots, F_\nu(X_d)) \sim C^t_{\Sigma,\nu}$. Because $\mathbf{X} \stackrel{d}{=} -\mathbf{X}$ and $F_\nu(-x) = 1 - F_\nu(x)$,

$$\begin{aligned}(U_1, \ldots, U_d) &\stackrel{d}{=} (F_\nu(-X_1), \ldots, F_\nu(-X_d)) \\ &= (1 - F_\nu(X_1), \ldots, 1 - F_\nu(X_d)) \\ &\stackrel{d}{=} (1 - U_1, \ldots, 1 - U_d).\end{aligned}$$

In other words, the Student copula is radially symmetric.

A.2. Proof of Lemma 1

Since $\widetilde{F}_\nu(x) = \mathbb{P}(X_j^2 \le x) = 2F_\nu(\sqrt{x}) - 1$, the Fisher copula corresponds to the joint distribution of $(\widetilde{F}_\nu(X_1^2), \ldots, \widetilde{F}_\nu(X_d^2)) = (2F_\nu(|X_1|) - 1, \ldots, 2F_\nu(|X_d|) - 1)$. Hence, using the fact that $-F_\nu^{-1}(x) = F_\nu^{-1}(1-x)$ because F_ν is symmetric around zero,

$$\begin{aligned}C^{\mathrm{F}}_{\Sigma,\nu}(u_1, \ldots, u_d) &= \mathbb{P}\{2F_\nu(|X_1|) - 1 \le u_1, \ldots, 2F_\nu(|X_d|) - 1 \le u_d\} \\ &= \mathbb{P}\left\{|X_1| \le F_\nu^{-1}\left(\frac{1+u_1}{2}\right), \ldots, |X_d| \le F_\nu^{-1}\left(\frac{1+u_d}{2}\right)\right\} \\ &= \mathbb{P}\left\{-F_\nu^{-1}\left(\frac{1+u_1}{2}\right) \le X_1 \le F_\nu^{-1}\left(\frac{1+u_1}{2}\right), \ldots, -F_\nu^{-1}\left(\frac{1+u_d}{2}\right) \le X_d \le F_\nu^{-1}\left(\frac{1+u_d}{2}\right)\right\} \\ &= \mathbb{P}\left\{F_\nu^{-1}\left(\frac{1-u_1}{2}\right) \le X_1 \le F_\nu^{-1}\left(\frac{1+u_1}{2}\right), \ldots, F_\nu^{-1}\left(\frac{1-u_d}{2}\right) \le X_d \le F_\nu^{-1}\left(\frac{1+u_d}{2}\right)\right\} \\ &= \mathbb{P}\left\{\frac{1-u_1}{2} \le F_\nu(X_1) \le \frac{1+u_1}{2}, \ldots, \frac{1-u_d}{2} \le F_\nu(X_d) \le \frac{1+u_d}{2}\right\}.\end{aligned}$$

Since by construction $(X_1, \ldots, X_d)$ follows a d-variate Student distribution, the vector $(U_1, \ldots, U_d) = (F_\nu(X_1), \ldots, F_\nu(X_d))$ is distributed from the Student copula. Therefore, one can write

$$C^{\mathrm{F}}_{\Sigma,\nu}(u_1, \ldots, u_d) = \mathbb{P}\left(\bigcap_{j=1}^d \left\{\frac{1-u_j}{2} \le U_j \le \frac{1+u_j}{2}\right\}\right). \tag{A.1}$$

Now it will be shown by induction that for an arbitrary continuous random vector $\mathbf{X} = (X_1, \ldots, X_d)$, one has for $a_1 < b_1, \ldots, a_d < b_d$ that

$$\mathbb{P}\left(\bigcap_{j=1}^d \{a_j \le X_j \le b_j\}\right) = \sum_{\boldsymbol{\epsilon} \in \{-1,1\}^d} \left(\prod_{j=1}^d \epsilon_j\right) \mathbb{P}\left(\bigcap_{j=1}^d \left\{X_j \le a_j^{\frac{1-\epsilon_j}{2}} \times b_j^{\frac{1+\epsilon_j}{2}}\right\}\right). \tag{A.2}$$

First note that the result is true when $d = 1$, since the right-hand side of (A.2) reduces to

$$\sum_{\epsilon_1 \in \{-1,1\}} \epsilon_1 \, \mathbb{P}\left(\left\{X_1 \le a_1^{\frac{1-\epsilon_1}{2}} \times b_1^{\frac{1+\epsilon_1}{2}}\right\}\right) = \mathbb{P}(X_1 \le b_1) - \mathbb{P}(X_1 \le a_1) = \mathbb{P}(a_1 \le X_1 \le b_1).$$

Now assuming that (A.2) is true for some $d \in \mathbb{N}$, one has

$$\begin{aligned}\mathbb{P}\left(\bigcap_{j=1}^{d+1} \{a_j \le X_j \le b_j\}\right) &= \mathbb{P}\left(\bigcap_{j=1}^{d} \{a_j \le X_j \le b_j\}, a_{d+1} \le X_{d+1} \le b_{d+1}\right) \\ &= \sum_{\boldsymbol{\epsilon} \in \{-1,1\}^d} \left(\prod_{j=1}^d \epsilon_j\right) \mathbb{P}\left(\bigcap_{j=1}^d \left\{X_j \le a_j^{\frac{1-\epsilon_j}{2}} \times b_j^{\frac{1+\epsilon_j}{2}}\right\}, a_{d+1} \le X_{d+1} \le b_{d+1}\right) \\ &= \sum_{\boldsymbol{\epsilon} \in \{-1,1\}^d} \left(\prod_{j=1}^d \epsilon_j\right) \left[\mathbb{P}\left(\bigcap_{j=1}^d \left\{X_j \le a_j^{\frac{1-\epsilon_j}{2}} \times b_j^{\frac{1+\epsilon_j}{2}}\right\}, X_{d+1} \le b_{d+1}\right)\right. \\ &\qquad \left. - \mathbb{P}\left(\bigcap_{j=1}^d \left\{X_j \le a_j^{\frac{1-\epsilon_j}{2}} \times b_j^{\frac{1+\epsilon_j}{2}}\right\}, X_{d+1} \le a_{d+1}\right)\right] \\ &= \sum_{\boldsymbol{\epsilon} \in \{-1,1\}^{d+1}} \left(\prod_{j=1}^{d+1} \epsilon_j\right) \mathbb{P}\left(\bigcap_{j=1}^{d+1} \left\{X_j \le a_j^{\frac{1-\epsilon_j}{2}} \times b_j^{\frac{1+\epsilon_j}{2}}\right\}\right).\end{aligned}$$

Hence, the formula in (A.2) holds for all $d \in \mathbb{N}$. Finally, letting $a_j = (1-u_j)/2$ and $b_j = (1+u_j)/2$ for all $j \in \{1,\ldots,d\}$ and upon noting that

$$a_j^{\frac{1-\epsilon_j}{2}} \times b_j^{\frac{1+\epsilon_j}{2}} = \left(\frac{1-u_j}{2}\right)^{\frac{1-\epsilon_j}{2}} \left(\frac{1+u_j}{2}\right)^{\frac{1+\epsilon_j}{2}} = \frac{1+\epsilon_j\, u_j}{2}\,,$$

one deduces from (A.1) and (A.2) that

$$\begin{aligned} C^{\mathrm{F}}_{\Sigma,\nu}(u_1,\ldots,u_d) &= \sum_{\boldsymbol{\epsilon}\in\{-1,1\}^d} \left(\prod_{j=1}^d \epsilon_j\right) \mathbb{P}\left(\bigcap_{j=1}^d \left\{U_j \le \frac{1+\epsilon_j\, u_j}{2}\right\}\right) \\ &= \sum_{\boldsymbol{\epsilon}\in\{-1,1\}^d} \left(\prod_{j=1}^d \epsilon_j\right) C^{\mathrm{t}}_{\Sigma,\nu}\left(\frac{1+\epsilon_1 u_1}{2},\ldots,\frac{1+\epsilon_d u_d}{2}\right). \end{aligned}$$

The expression for the density follows easily from

$$\begin{aligned} c^{\mathrm{F}}_{\Sigma,\nu}(u_1,\ldots,u_d) &= \frac{\partial^d}{\partial u_1\cdots\partial u_d} C^{\mathrm{F}}_{\Sigma,\nu}(u_1,\ldots,u_d) \\ &= \sum_{\boldsymbol{\epsilon}\in\{-1,1\}^d} \left(\prod_{j=1}^d \epsilon_j\right) \frac{\partial^d}{\partial u_1\cdots\partial u_d} C^{\mathrm{t}}_{\Sigma,\nu}\left(\frac{1+\epsilon_1 u_1}{2},\ldots,\frac{1+\epsilon_d u_d}{2}\right) \\ &= \sum_{\boldsymbol{\epsilon}\in\{-1,1\}^d} \left(\prod_{j=1}^d \epsilon_j\right) \left(\prod_{j=1}^d \frac{\epsilon_j}{2}\right) c^{\mathrm{t}}_{\Sigma,\nu}\left(\frac{1+\epsilon_1 u_1}{2},\ldots,\frac{1+\epsilon_d u_d}{2}\right) \\ &= \frac{1}{2^d} \sum_{\boldsymbol{\epsilon}\in\{-1,1\}^d} c^{\mathrm{t}}_{\Sigma,\nu}\left(\frac{1+\epsilon_1 u_1}{2},\ldots,\frac{1+\epsilon_d u_d}{2}\right). \end{aligned}$$

A.3. Proof of Proposition 1

A.3.1. Upper tail dependence coefficient

Let (X_1, X_2) follow the Student distribution $T_{\rho,\nu}$ with parameter ρ and ν degrees of freedom. Because the cdf of X_ℓ^2 is $F_\ell(x) = 2\,F_\nu(\sqrt{x}) - 1$, $\ell = 1, 2$, one has by definition that

$$\begin{aligned} \lambda_U(C^{\mathrm{F}}_{\rho,\nu}) &= \lim_{u\uparrow 1} \mathbb{P}\left\{X_1^2 > F_1^{-1}(u) \middle| X_2^2 > F_2^{-1}(u)\right\} \\ &= \lim_{u\uparrow 1} \mathbb{P}\left\{X_1^2 > \left(F_\nu^{-1}\left(\frac{u+1}{2}\right)\right)^2 \middle| X_2^2 > \left(F_\nu^{-1}\left(\frac{u+1}{2}\right)\right)^2\right\} \\ &= \lim_{s\uparrow 1} \mathbb{P}\left\{X_1^2 > \left(F_\nu^{-1}(s)\right)^2 \middle| X_2^2 > \left(F_\nu^{-1}(s)\right)^2\right\} \\ &= \lim_{s\uparrow 1} \mathbb{P}\left\{|X_1| > F_\nu^{-1}(s) \middle| |X_2| > F_\nu^{-1}(s)\right\} \\ &= \lim_{s\uparrow 1} \mathbb{P}\left\{X_1 > F_\nu^{-1}(s) \middle| |X_2| > F_\nu^{-1}(s)\right\} + \mathbb{P}\left\{-X_1 > F_\nu^{-1}(s) \middle| |X_2| > F_\nu^{-1}(s)\right\}. \end{aligned} \tag{A.3}$$

For the first summand in the last expression, note that

$$\begin{aligned} &\mathbb{P}\left\{X_1 > F_\nu^{-1}(s) \middle| |X_2| > F_\nu^{-1}(s)\right\} \\ &= \frac{\mathbb{P}\left\{X_1 > F_\nu^{-1}(s), |X_2| > F_\nu^{-1}(s)\right\}}{\mathbb{P}\left\{|X_2| > F_\nu^{-1}(s)\right\}} \\ &= \frac{\mathbb{P}\left\{X_1 > F_\nu^{-1}(s), X_2 > F_\nu^{-1}(s)\right\} + \mathbb{P}\left\{X_1 > F_\nu^{-1}(s), -X_2 > F_\nu^{-1}(s)\right\}}{\mathbb{P}\left\{|X_2| > F_\nu^{-1}(s)\right\}} \\ &= \frac{\mathbb{P}\left\{X_2 > F_\nu^{-1}(s)\right\}}{\mathbb{P}\left\{|X_2| > F_\nu^{-1}(s)\right\}} \mathbb{P}\left\{X_1 > F_\nu^{-1}(s) \middle| X_2 > F_\nu^{-1}(s)\right\} \\ &\quad + \frac{\mathbb{P}\left\{-X_2 > F_\nu^{-1}(s)\right\}}{\mathbb{P}\left\{|X_2| > F_\nu^{-1}(s)\right\}} \mathbb{P}\left\{X_1 > F_\nu^{-1}(s) \middle| -X_2 > F_\nu^{-1}(s)\right\} \\ &= \frac{1}{2}\left[\mathbb{P}\left\{X_1 > F_\nu^{-1}(s) \middle| X_2 > F_\nu^{-1}(s)\right\} + \mathbb{P}\left\{X_1 > F_\nu^{-1}(s) \middle| -X_2 > F_\nu^{-1}(s)\right\}\right], \end{aligned}$$

where the last equality follows from the fact that for any $a > 0$,

$$\frac{\mathbb{P}(-X_2 > a)}{\mathbb{P}(|X_2| > a)} = \frac{\mathbb{P}(X_2 > a)}{\mathbb{P}(|X_2| > a)} = \frac{1 - F_\nu(a)}{2\{1 - F_\nu(a)\}} = \frac{1}{2}\,.$$

Similarly, for the second summand, one can write

$$\begin{aligned}
&\mathbb{P}\left\{-X_1 > F_\nu^{-1}(s) \middle| \, |X_2| > F_\nu^{-1}(s)\right\} \\
&\quad = \frac{1}{2}\left[\mathbb{P}\left\{-X_1 > F_\nu^{-1}(s) \middle| X_2 > F_\nu^{-1}(s)\right\} + \mathbb{P}\left\{-X_1 > F_\nu^{-1}(s) \middle| -X_2 > F_\nu^{-1}(s)\right\}\right].
\end{aligned}$$

Putting these two expressions in (A.3) yields

$$\begin{aligned}
\lambda_U(C^{\mathrm{F}}_{\rho,\nu}) &= \frac{1}{2}\lim_{s\uparrow 1}\left[\mathbb{P}\left\{X_1 > F_\nu^{-1}(s) \middle| X_2 > F_\nu^{-1}(s)\right\} + \mathbb{P}\left\{X_1 > F_\nu^{-1}(s) \middle| -X_2 > F_\nu^{-1}(s)\right\}\right] \\
&\quad + \frac{1}{2}\lim_{s\uparrow 1}\left[\mathbb{P}\left\{-X_1 > F_\nu^{-1}(s) \middle| X_2 > F_\nu^{-1}(s)\right\} + \mathbb{P}\left\{-X_1 > F_\nu^{-1}(s) \middle| -X_2 > F_\nu^{-1}(s)\right\}\right] \\
&= \lim_{s\uparrow 1}\left[\mathbb{P}\left\{X_1 > F_\nu^{-1}(s) \middle| X_2 > F_\nu^{-1}(s)\right\} + \mathbb{P}\left\{X_1 > F_\nu^{-1}(s) \middle| -X_2 > F_\nu^{-1}(s)\right\}\right] \\
&= \lambda_U(C^t_{\rho,\nu}) + \lambda_U(C^t_{-\rho,\nu}),
\end{aligned}$$

where the last two equalities follow from the fact that $(-X_1, -X_2) \stackrel{d}{=} (X_1, X_2) \sim T_{\rho,\nu}$ and $(-X_1, X_2) \stackrel{d}{=} (X_1, -X_2) \sim T_{-\rho,\nu}$.

A.3.2. Lower tail dependence coefficient

By definition, for $(X_1, X_2) \sim T_{\rho,\nu}$ and $F_\ell(x) = 2\,F_\nu(\sqrt{x}) - 1$, $\ell = 1, 2$,

$$\begin{aligned}
\lambda_L(C^{\mathrm{F}}_{\rho,\nu}) &= \lim_{u\downarrow 0}\mathbb{P}\left\{X_1^2 < F_1^{-1}(u) \middle| X_2^2 < F_2^{-1}(u)\right\} \\
&= \lim_{u\downarrow 0}\mathbb{P}\left\{X_1^2 < \left(F_\nu^{-1}\left(\frac{u+1}{2}\right)\right)^2 \middle| \, X_2^2 < \left(F_\nu^{-1}\left(\frac{u+1}{2}\right)\right)^2\right\} \\
&= \lim_{s\downarrow 0}\mathbb{P}\left(X_1^2 < s^2 \middle| X_2^2 < s^2\right).
\end{aligned}$$

If $H^{\mathrm{t}}_{\rho,\nu}$ is the joint distribution function of (X_1, X_2), one has for $s \geq 0$ that

$$\mathbb{P}\left(X_1^2 < s^2 \middle| X_2^2 < s^2\right) = \frac{H^{\mathrm{t}}_{\rho,\nu}(s,s) - H^{\mathrm{t}}_{\rho,\nu}(-s,s) - H^{\mathrm{t}}_{\rho,\nu}(s,-s) + H^{\mathrm{t}}_{\rho,\nu}(-s,-s)}{2F_\nu(s) - 1}.$$

Since $H^{\mathrm{t}}_{\rho,\nu}$ is continuous, in particular at $(x_1, x_2) = (0, 0)$,

$$\begin{aligned}
\lambda_L(C^{\mathrm{F}}_{\rho,\nu}) &= \lim_{s\downarrow 0}\left\{\frac{H^{\mathrm{t}}_{\rho,\nu}(s,s) - H^{\mathrm{t}}_{\rho,\nu}(-s,s) - H^{\mathrm{t}}_{\rho,\nu}(s,-s) + H^{\mathrm{t}}_{\rho,\nu}(-s,-s)}{2F_\nu(s) - 1}\right\} \\
&= \lim_{s\downarrow 0}\frac{H^{\mathrm{t}}_{\rho,\nu}(s,s)}{2F_\nu(s) - 1} - \lim_{s\downarrow 0}\frac{H^{\mathrm{t}}_{\rho,\nu}(-s,s)}{2F_\nu(s) - 1} - \lim_{s\downarrow 0}\frac{H^{\mathrm{t}}_{\rho,\nu}(s,-s)}{2F_\nu(s) - 1} + \lim_{s\downarrow 0}\frac{H^{\mathrm{t}}_{\rho,\nu}(-s,-s)}{2F_\nu(s) - 1} \\
&= \lim_{s\downarrow 0}\frac{H^{\mathrm{t}}_{\rho,\nu}(s,s)}{2F_\nu(s) - 1} - \lim_{s\downarrow 0}\frac{H^{\mathrm{t}}_{\rho,\nu}(s,s)}{2F_\nu(s) - 1} - \lim_{s\downarrow 0}\frac{H^{\mathrm{t}}_{\rho,\nu}(s,s)}{2F_\nu(s) - 1} + \lim_{s\downarrow 0}\frac{H^{\mathrm{t}}_{\rho,\nu}(s,s)}{2F_\nu(s) - 1} \\
&= 0.
\end{aligned}$$

A.4. Proof of Proposition 2

The proof is straightforward. If the d-variate Fisher copula was radially symmetric, then all the marginal bivariate copulas would be radially symmetric as well. As a consequence, one would have in particular that $\lambda_L(C^{\mathrm{F}}_{\Sigma_{jj'},\nu}) = \lambda_U(C^{\mathrm{F}}_{\Sigma_{jj'},\nu})$ for all $j < j' \in \{1, \ldots, d\}$. But as long as $\nu < \infty$, this contradicts Proposition 1 whatever the value of $\Sigma_{jj'}$; therefore the Fisher copula must be radially asymmetric.

A.5. Estimation of the tail dependence coefficients

Let $(X_{11}, X_{12}), \ldots, (X_{n1}, X_{n2})$ be independent copies of a pair (X_1, X_2). As proposed by Schmidt and Stadtmüller (2006), the lower and upper tail dependence coefficients $\lambda_L(X_1, X_2)$ and $\lambda_U(X_1, X_2)$ could be estimated using the empirical copula

$$C_n(u_1, u_2) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}\left(\frac{R_{i1}}{n} \leq u_1, \frac{R_{i2}}{n} \leq u_2\right),$$

where $(R_{11}, R_{12}), \ldots, (R_{n1}, R_{n2})$ is the sample of the pairs of ranks. Specifically, from the definitions of $\lambda_L(X_1, X_2)$ and $\lambda_U(X_1, X_2)$ in (1), one has for some threshold $\kappa > 0$ that

$$\widehat{\lambda}_L = \frac{C_n(\kappa, \kappa)}{\kappa} = \frac{1}{n\kappa}\sum_{i=1}^{n}\mathbb{I}\left(\frac{R_{i1}}{n} \leq \kappa, \frac{R_{i2}}{n} \leq \kappa\right)$$

and

$$\widehat{\lambda}_U = \frac{\bar{C}_n(\kappa,\kappa)}{\kappa} = \frac{1}{n\kappa}\sum_{i=1}^{n}\mathbb{I}\left(1-\frac{R_{i1}}{n}\leq\kappa, 1-\frac{R_{i2}}{n}\leq\kappa\right) \approx \frac{1}{n\kappa}\sum_{i=1}^{n}\mathbb{I}\left(\frac{R_{i1}}{n}>1-\kappa, \frac{R_{i2}}{n}>1-\kappa\right).$$

In the data analysis of Section 4, the threshold has been set to $\kappa = 0.02$.

REFERENCES

Aas K, Czado C, Frigessi A, Bakken H, 2009. Pair-copula constructions of multiple dependence. *Insurance Math. Econom.* **44**(2): 182–198.

Bàrdossy A, 2006. Copula-based geostatistical models for groundwater quality parameters. *Water Resources Research* **42**(11): 1–12.

Evin G, Favre AC, Hingray B, 2017. Stochastic generation of multi-site daily precipitation for the assessment of extreme floods in Switzerland. *Hydrology and Earth System Sciences Discussions* **2017**: 1–36.

Genest C, Ghoudi K, Rivest LP, 1995. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* **82**(3): 543–552.

Higham NJ, 2002. Computing the nearest correlation matrix—a problem from finance. *IMA J. Numer. Anal.* **22**(3): 329–343.

Joe H, 2015. *Dependence modeling with copulas*, volume 134 of *Monographs on Statistics and Applied Probability*. CRC Press, Boca Raton, FL, xviii+462 pp.

Lehmann EA, Phatak A, Stephenson A, Lau R, 2016. Spatial modelling framework for the characterisation of rainfall extremes at different durations and under climate change. *Environmetrics* **27**(4): 239–251, env.2389.

Mai JF, Scherer M, 2012. *Simulating copulas*, volume 4 of *Series in Quantitative Finance*. Imperial College Press, London, xiv+295 pp., stochastic models, sampling algorithms, and applications.

McNeil AJ, Frey R, Embrechts P, 2005. *Quantitative risk management: Concepts, techniques and tools*. Princeton University Press, Princeton, NJ.

Müller D, Czado C, 2017. Dependence modeling in ultra high dimensions with vine copulas and the graphical lasso. *arXiv preprint arXiv:1709.05119* .

Nelsen RB, 2006. *An introduction to copulas*. Springer Series in Statistics, Springer, New York, second edition, xiv+269 pp.

Quessy JF, Rivest LP, Toupin MH, 2016. On the family of multivariate chi-square copulas. *J. Multivariate Anal.* **152**: 40–60.

Schmidt R, Stadtmüller U, 2006. Non-parametric estimation of tail dependence. *Scand. J. Statist.* **33**(2): 307–335.

Serinaldi F, Bárdossy A, Kilsby CG, 2015. Upper tail dependence in rainfall extremes: would we know it if we saw it? *Stochastic Environmental Research and Risk Assessment* **29**(4): 1211–1233.

Sun Y, Stein ML, 2015. A stochastic space-time model for intermittent precipitation occurrences. *Ann. Appl. Stat.* **9**(4): 2110–2132.

Varin C, Reid N, Firth D, 2011. An overview of composite likelihood methods. *Statist. Sinica* **21**(1): 5–42.

10.3.4 Student versus Fisher copulas: results for 1000yrs simulations

There are four models in competition from the combination of precipitation two disaggregation schemes and two copulas:

- '1D' indicates that the fit is done directly on 1-day precipitation while '3D' means that the fit is done on 3-day precipitation with then a temporal disaggregation at daily scale (Evin et al., 2018a).
- 'Student' or 'Fisher' (Favre et al., 2018) indicates the copula used in the model.

This comparison was done on 100 scenarios of 100 years for each model by using performance categorization criteria used by Bennett et al. (2017) (Table A. 1)

Table A. 1 Performance categorization criteria from Bennett et al. (2017).

Performance Classification	Key	Test
'good'		Observed metric inside 90% limits (case 1)
'fair'		Observed metric outside 90% limits but within the 99.7% limits (case 2) OR absolute relative difference between the observed metric and the average simulated metrics is 5% or less (case 3)
'poor'		Otherwise (case 4)

The models that were fitted on 1-day precipitation are slightly better concerning simulation of the number of wet days (Figure A. 1 At site number of wet days for all sites and months: inter-annual and standard deviation. The 90% probability limits are shown for the different seasons. Overall performance is represented by the percentage of 'good', 'fair' and 'poor' performance for all sites and months.(Figure A. 1), the distribution of the length of dry (Figure A. 2) and wet spells (Figure A. 3). The performance between the four models are very similar regarding daily amounts (Figure A. 4) and 1-day annual maxima (Figure A. 5). However, the models fitted on 3-day precipitation clearly outperform those fitted on 1-day precipitation for the simulation of 3-day extreme events (Figure A. 6). With respect to latter characteristic the GWEX-3D-Student and the GWEX-3D-Fisher models appear to be the most suitable for hydrological purpose for which 3-day extreme events are crucial. The choice of the copula affects the results only little and both copulas can be retained.

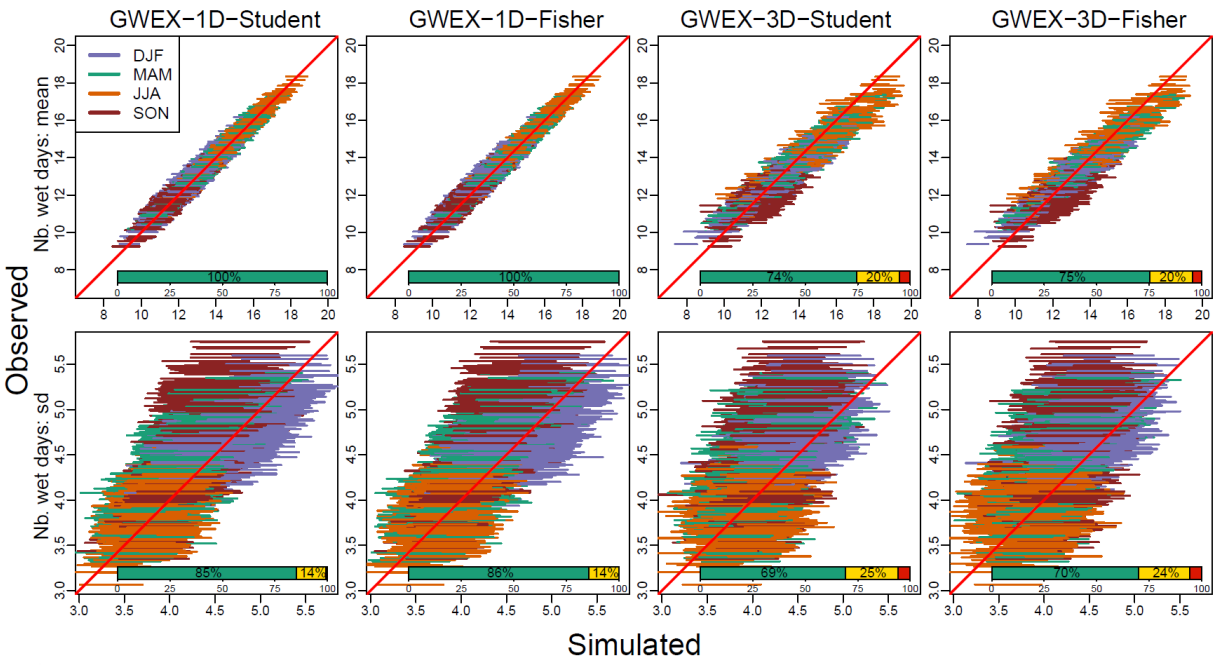


Figure A. 1 At site number of wet days for all sites and months: inter-annual and standard deviation. The 90% probability limits are shown for the different seasons. Overall performance is represented by the percentage of 'good', 'fair' and 'poor' performance for all sites and months.

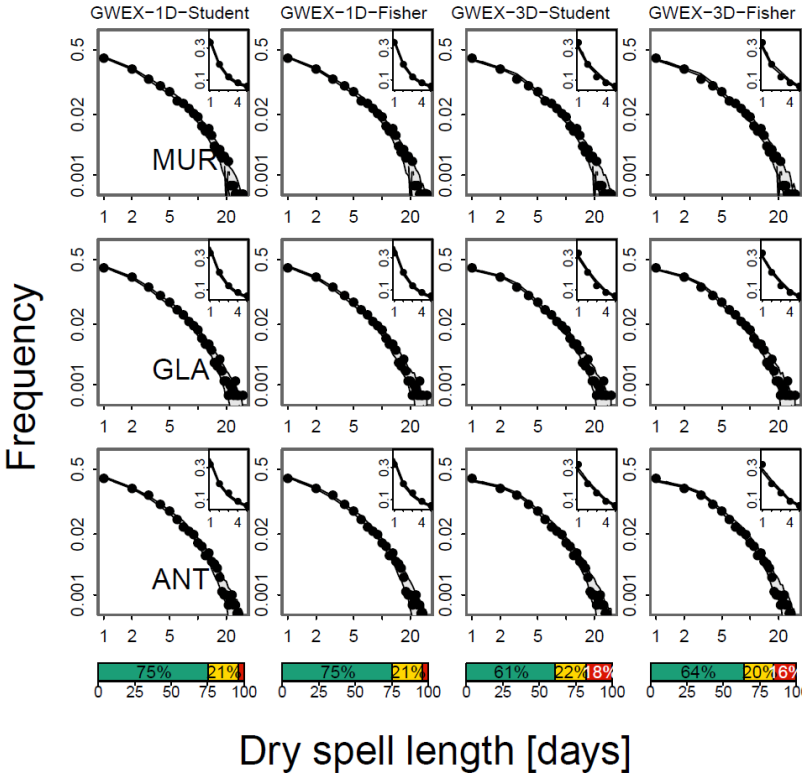


Figure A. 2 Distribution of dry spell lengths at the stations. The 90% probability limits are shown. Overall performance is represented by the percentage of 'good', 'fair' and 'poor' performance for all sites. Inset plots provide a zoom for durations of 1 to 5 days.

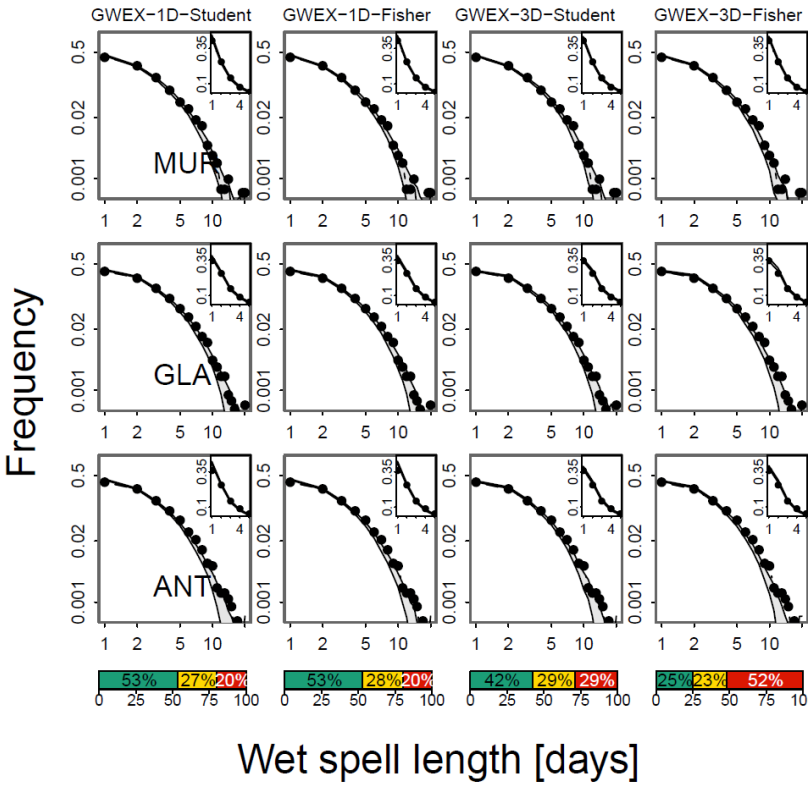


Figure A. 3 Distribution of wet spell lengths at the stations. The 90% probability limits are shown. Overall performance is represented by the indicated percentage of 'good', 'fair' and 'poor' performance for all sites. Inset plots provide a zoom for durations of 1 to 5 days.

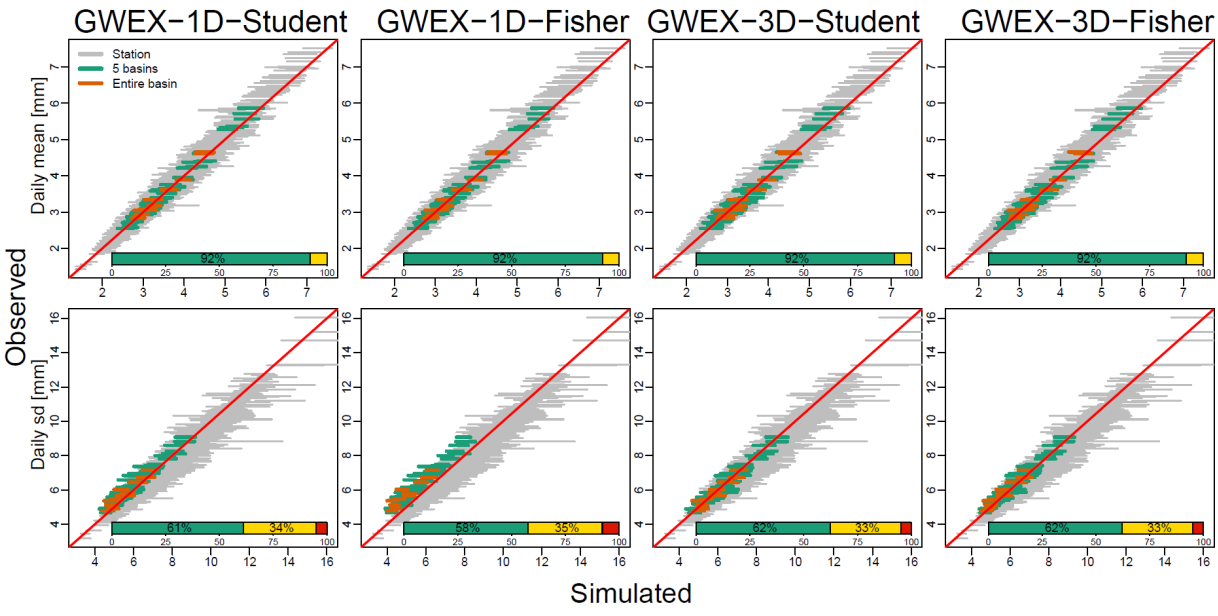


Figure A. 4 Daily amounts or all spatial scales and months: inter-annual mean (top) and standard deviation (bottom). The 90% probability limits are shown. Overall performance is represented by the percentage of 'good', 'fair' and 'poor' performance for all spatial scales and months.

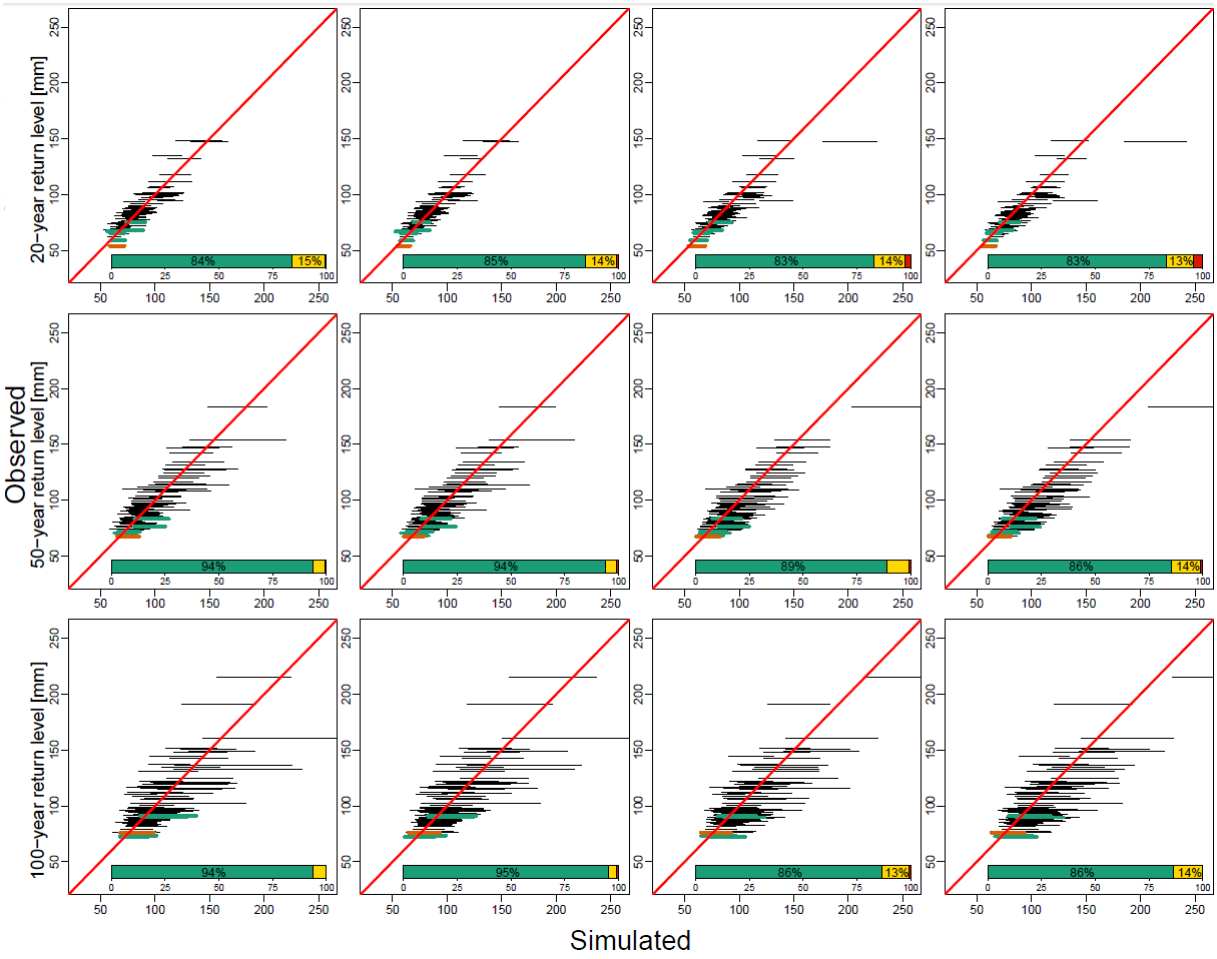


Figure A. 5 Daily annual maxima for all spatial scales: 20-year (top plots), 50-year (middle plots) and 100-years (bottom plots) return periods. The 90% probability limits are shown. The overall performance is represented by the percentage of 'good', 'fair' and 'poor' performance for all spatial scales.

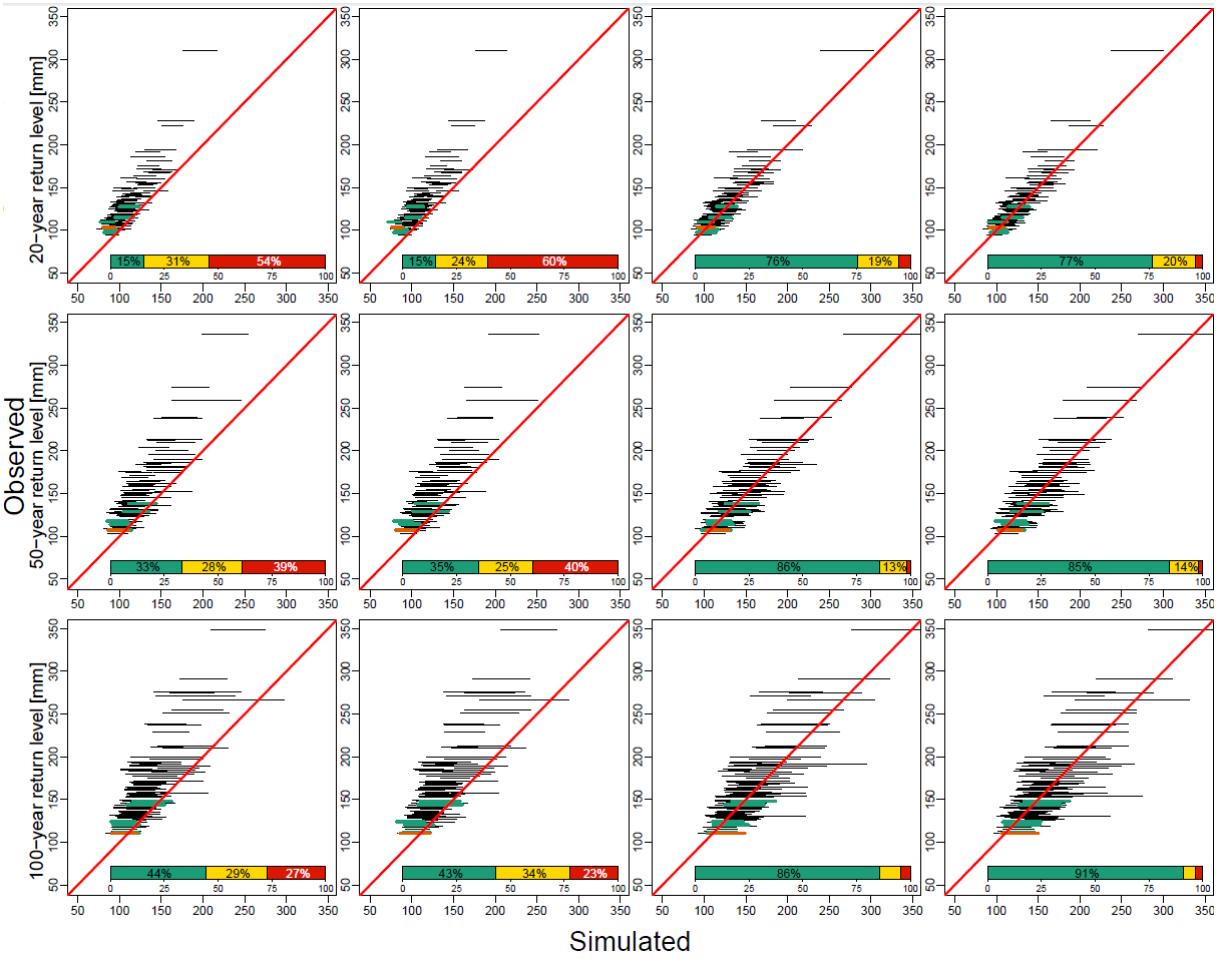


Figure A. 6 3-day annual maxima for all spatial scales: 20-year (top plots), 50-year (middle plots) and 100-years (bottom plots) return periods. The 90% probability limits are shown. Overall performance is represented by the indicated percentage of 'good', 'fair' and 'poor' performance for all spatial scales.

10.3.5 Sensitivity analysis on the extrapolation of the shape parameter of the E-GPD

EXAR

Working report

August 26, 2019

Sensitivity analysis on the parameter ξ of the EGPD model

Gilles Nicolet, Damien Raynaud, Guillaume Evin, Jérémy Chardon, Anne-Catherine Favre (Pr.) and Benoit Hingray (Res. Sc.)

1 Introduction

The first step to fit the GWEX model is the estimation of the shape parameter (noted ξ) of the extended Genalized Pareto Distribution (E-GPD, Papastathopoulos and Tawn (2013)). For the estimation of the GWEX, we have selected a dataset of 105 stations (called here the GWEX stations) for the period from 1930 to 2015 from a larger dataset (Figure 1(a)). The availability of precipitation observations is very different depending on the years (Figure 1(b)). Howevern the estimation of the shape parameter of the distributions used in extreme value statistic is a very tricky problem whose a poor estimation may have major consequences. That is why for the step of the ξ estimation, we use all the data at our disposal, that is to say 666 stations from 1864 to 2015. The ξ estimation is done by a regional approach (Evin et al., 2016) for each season (DJF, MAM, JJA and SON). The importance and the difficulty to estimate the shape parameter is also the reason why the estimation is done seasonally, whereas it is done monthly for the other steps of the GWEX fitting. This importance and this difficulty to deal with this issue is our main motivation to investigate here the sensitivity in the estimation of the parameter ξ in the GWEX fitting and its impact in the estimation of high return levels.

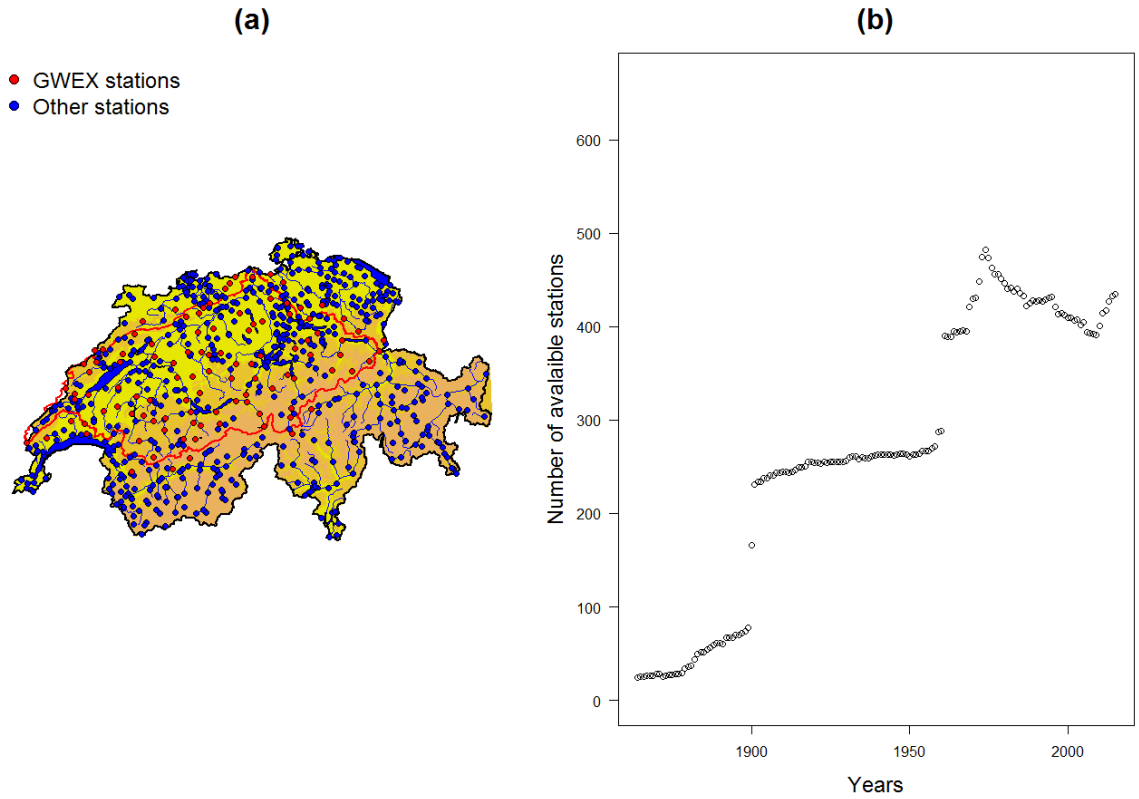


Figure 1: (a) The 666 stations used in the shape parameter ξ estimation. The 105 selected stations used in all the other steps of GWEX fitting (the GWEX stations) are in red. (b) The number of available stations from 1864 to 2015.

2 Bootstrap procedure

As the shape parameter estimation in the GWEX fitting, the bootstrap procedure is applied using all the data at our disposal (666 stations from 1864 to 2015), including stations and years with few observations. The bootstrap procedure is described as follow:

1. 152 years are randomly chosen with replacement among the period from 1864 to 2015; in order to approximately keep the same number of observation in the bootstrap dataset, a probability weight proportional to the number of available stations (Figure 1(b)) is given to each year in the random selection;

2. the regional estimation of the ξ parameter used in the GWEX fitting is applied to this new dataset;

3. this procedure is replicated until we have 30 new estimations of the parameter ξ for the 666 stations.

A part of the obtained ξ maps are shown in the Figure 2 and the others in the appendix B. We may observe that, even if some maps are very different from the original maps (appendix A), most of them are qualitatively similar.

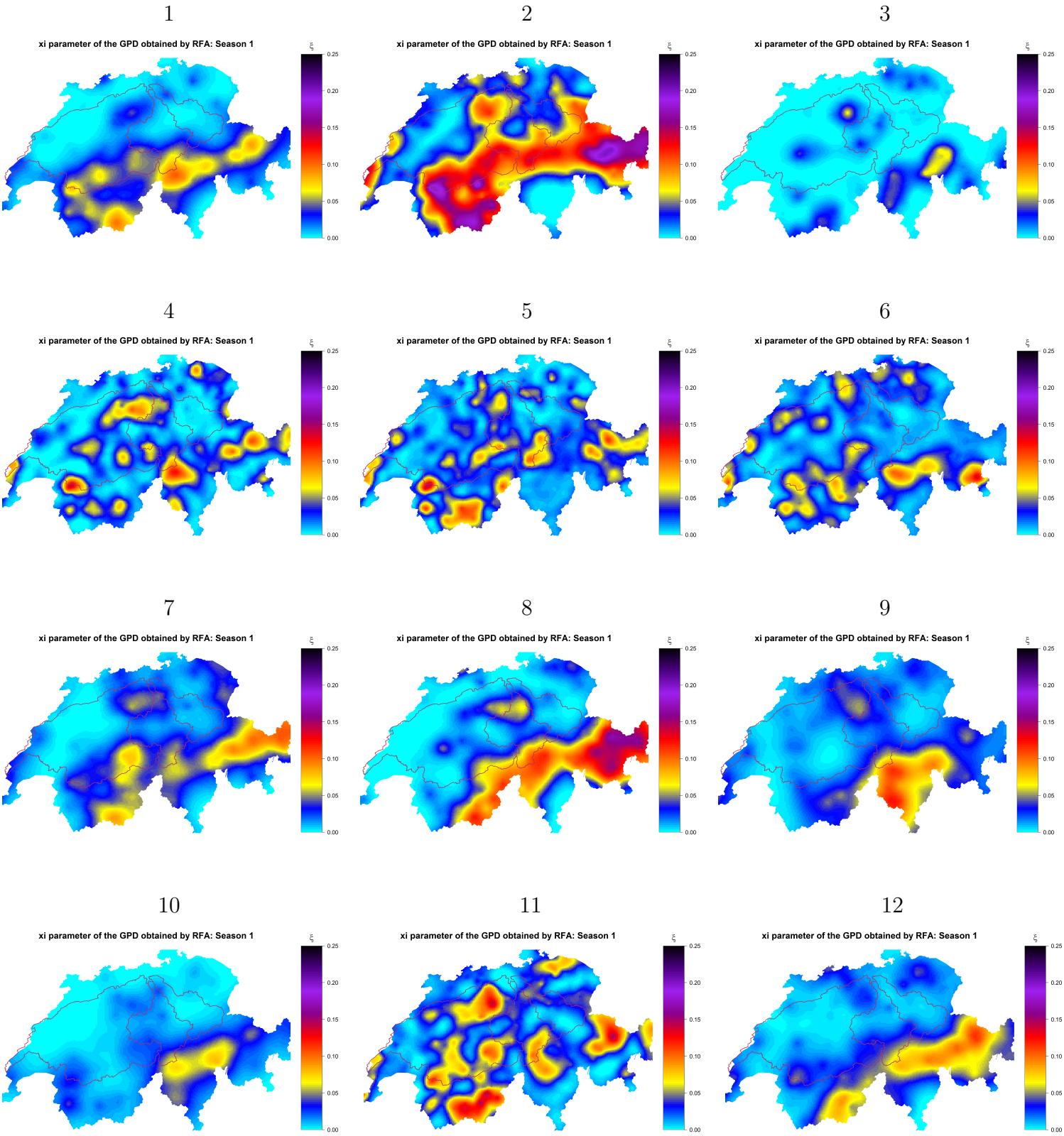


Figure 2: Maps of bootstrap ξ estimates (Winter, sample 1 to sample 12). The other maps are shown in the appendix B.

3 Effect of the sensitivity of the parameter ξ on the estimation of the 10'000-year return levels

In order to assess the impact of the sensitivity of the estimation of the parameter ξ in the computation of high return levels, we estimate, for each of the 30 boostrap datasets and at each station the two remaining parameters of the E-GPD, the scale parameter σ and the parameter κ of the transformation, by using the probability weighted moment method (Naveau et al., 2016) at each station (similarly as that is done in the GWEX estimation). In order to reproduce the procedure done in the GWEX fitting, these estimations are made monthly. Then, we use these estimates of the E-GPD at each station and for the 30+1 datasets to compute the 10'000-year return levels for the 105 GWEX stations.

Figure 3 and 4 present the boxplots of the ratios of these 10'000-year return levels relatively to the those estimated using the entire dataset. These boxplots are very similar in a same season, due to the fact that the shape parameter ξ is estimated for each season and the two other E-GPD parameters are estimated conditionally to ξ, confirming the strong impact of this parameter on the estimation of high return levels.

The deviation of ratios is rather limited because the seasonal first and third quartiles are all between 90% and 120% (Table 1), showing that most of ratios range between these two values. In addition, the seasonal medians are close to 100% (between 96% and 104%), indicating a absence of bias in the original estimation of the return levels. However, one may observe very large ratios in the Spring with 52 ratios above 200% and 541 ratios (i.e. 5.7% of the total number of ratios in the Spring) above 150%.

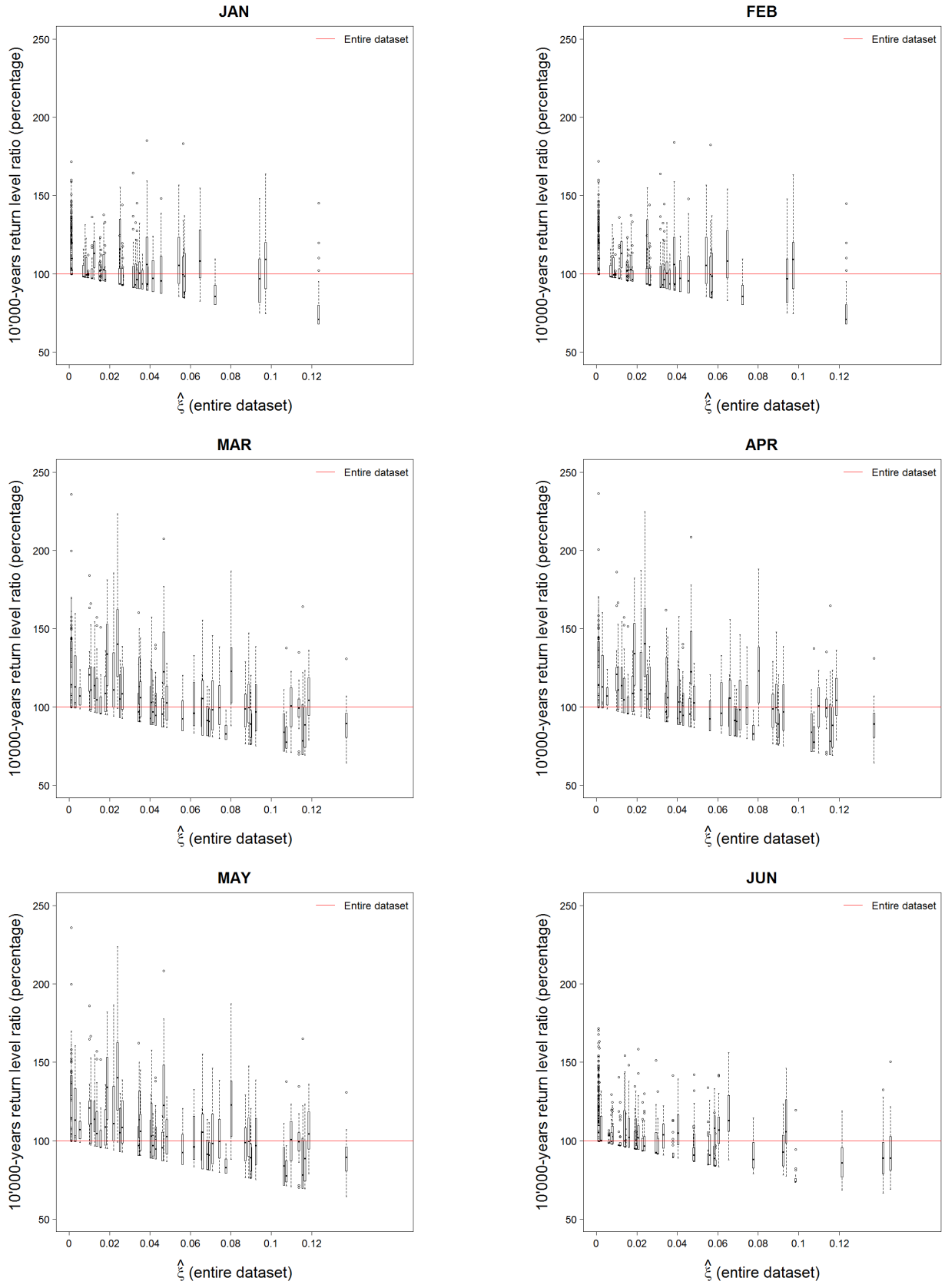


Figure 3: Boxplots of the 30 bootstrap 10'000 year return level ratios relatively to those estimated using the entire dataset for the 105 selected stations (from January to June).

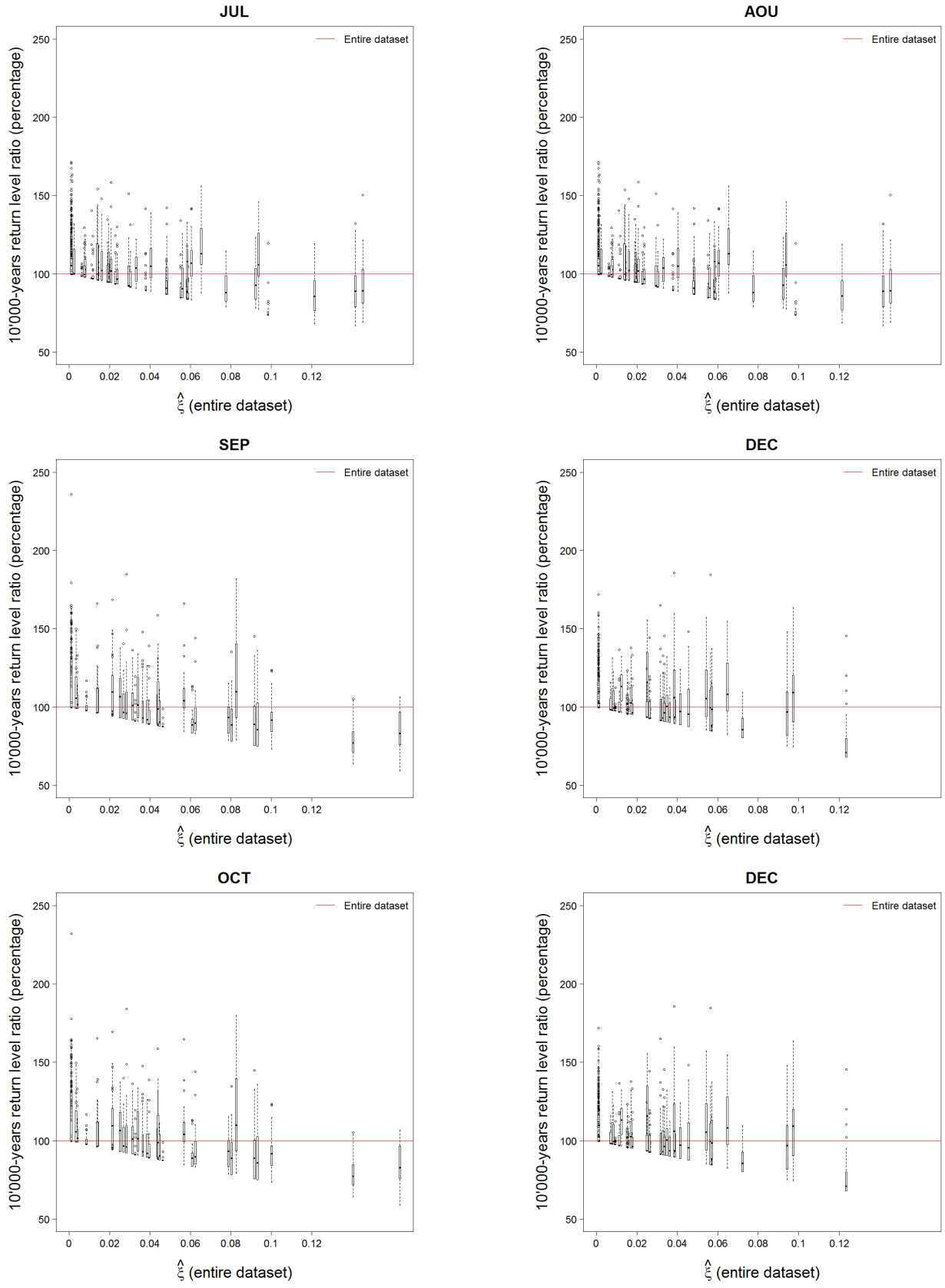


Figure 4: Boxplots of the 30 bootstrap 10'000 year return level ratios relatively to those estimated using the entire dataset for the 105 selected stations (from July to December).

	Winter	Spring	Summer	Autumn
Max	185.63	304.15	181.74	244.89
Q3	100	119.25	110.59	105.63
Median	96.47	103.42	102.36	100
Q1	90.39	97.26	99.63	90.82
Min	62.20	57.44	63.20	47.11

Table 1: Maximum, third quartile, median, first quartile and minimum for each season. The three months of the season, the 105 stations and the 30 replications are considered together (i.e., 9450 ratios considered for each season).

	Winter	Spring	Summer	Autumn
≥ 200	0	52	0	7
$[150, 200)$	43	489	65	127
$[125, 150)$	275	1267	554	504
$[75, 125)$	9005	7387	8782	8589
$[50, 75)$	127	255	49	219
< 50	0	0	0	4

Table 2: Number of ratios in each interval for each season. The three months of the season, the 105 stations and the 30 replications are considered together (i.e., 9450 ratios considered for each season).

4 Conclusion

The results of this bootstrap procedure show that the estimation of the parameter ξ can be considered reasonably robust comparing to the difficulty of the task, 89% of the return level estimates show a deviation less than 25%, while 4% only show a deviation more than 50%. In addition, one should precise that, as GWEX provides areal values of precipitation at the end of the procedure, the influence of potential very large pointwise simulated precipitations is mitigated by the spatial aggregation step.
However, the differences which appear in the bootstrap ξ maps and the large ratios of return levels in Spring suggest that the study of the possibility of more robust procedures could be very useful.

References

Evin, G., Blanchet, J., Paquet, E., Garavaglia, F., and Penot, D. (2016). A regional model for extreme rainfall based on weather patterns subsampling. *Journal of Hydrology*, 541:1185–1198.

Naveau, P., Huser, R., Ribereau, P., and Hannart, A. (2016). Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. *Water Resources Research*, 52(4):2753–2769.

Papastathopoulos, I. and Tawn, J. A. (2013). Extended generalised Pareto models for tail estimation. *Journal of Statistical Planning and Inference*, 143(1):131–143.

A Original ξ maps

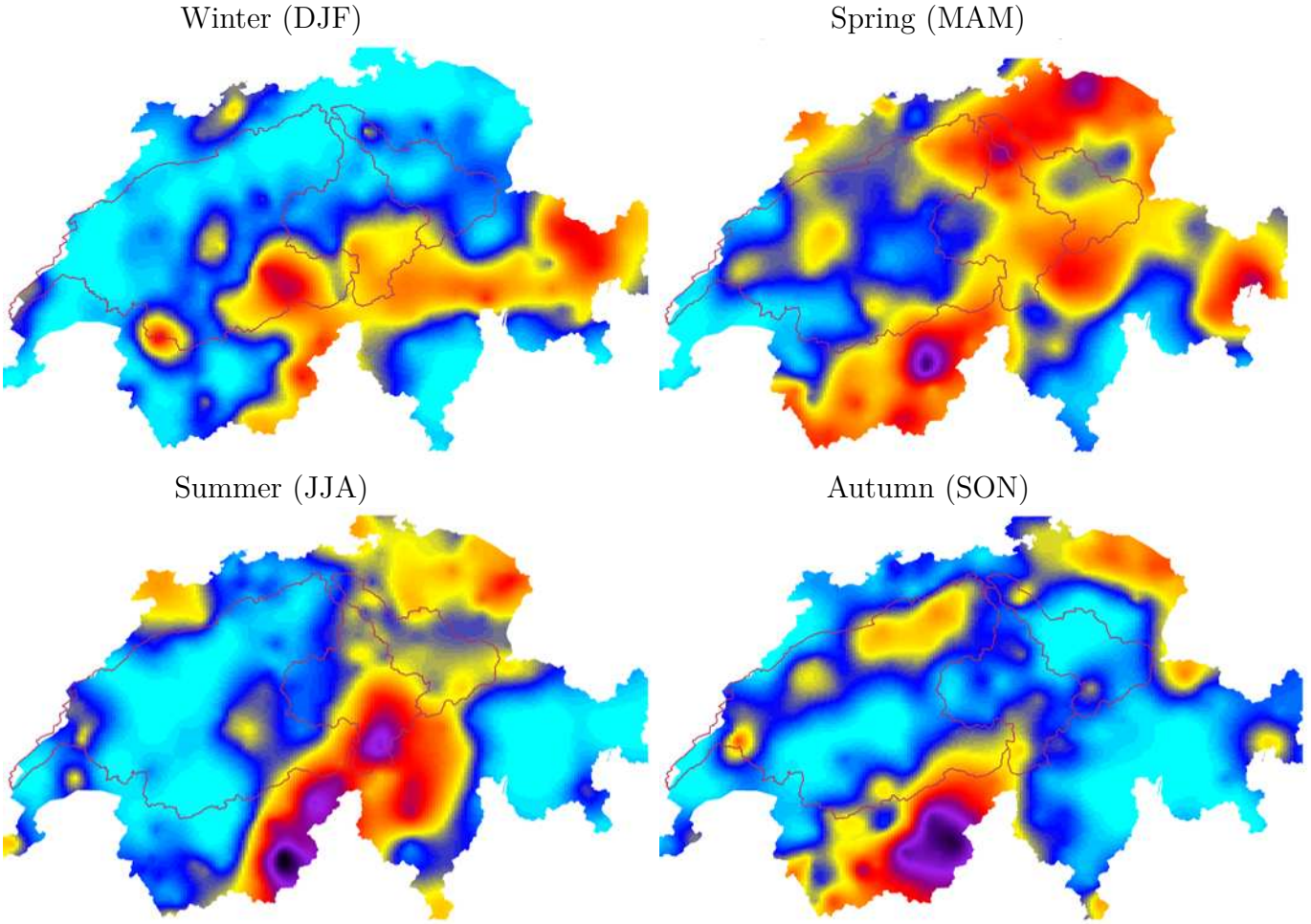
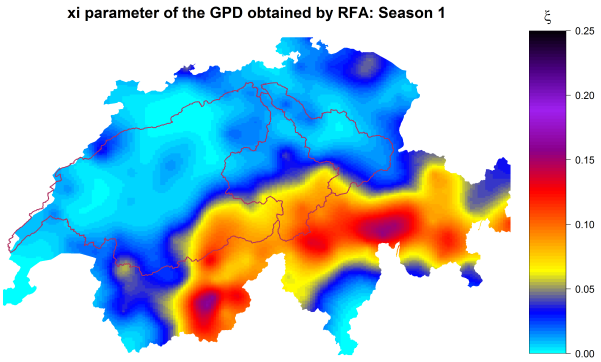


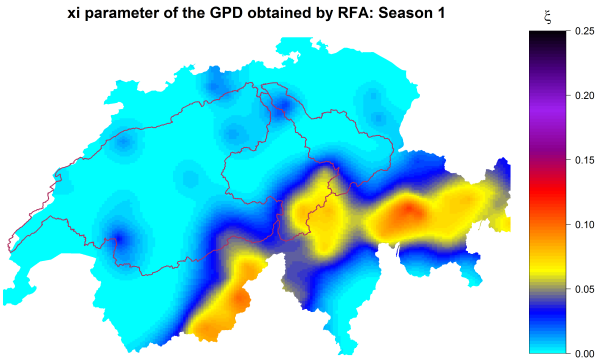
Figure 5: Original maps of bootstrap ξ estimates for Winter (DJF), Spring (MAM), Summer (JJA) and Autumn (SON).

B Boostrap ξ maps

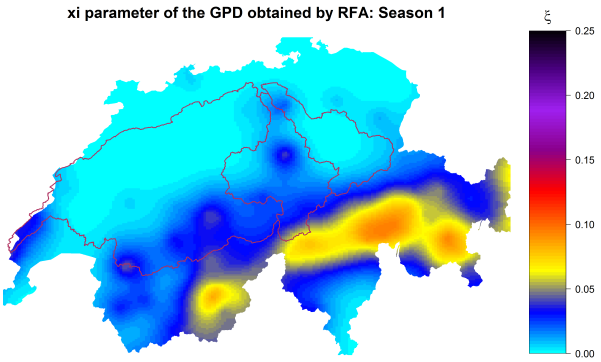
13



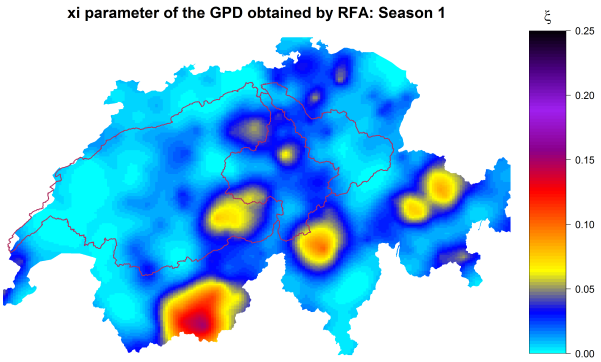
14



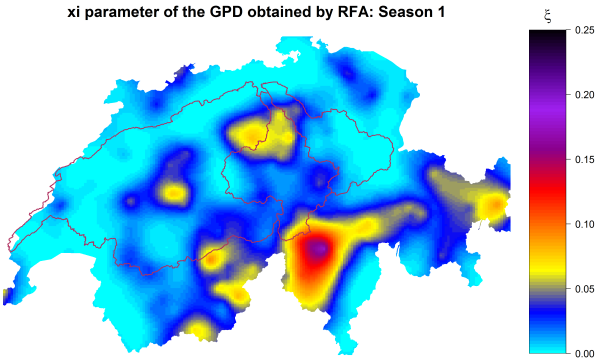
15



16



17



18

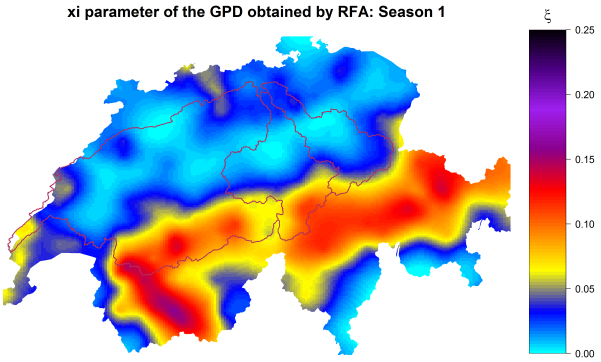


Figure 6: Maps of bootstrap ξ estimates (Winter, sample 13 to sample 18)

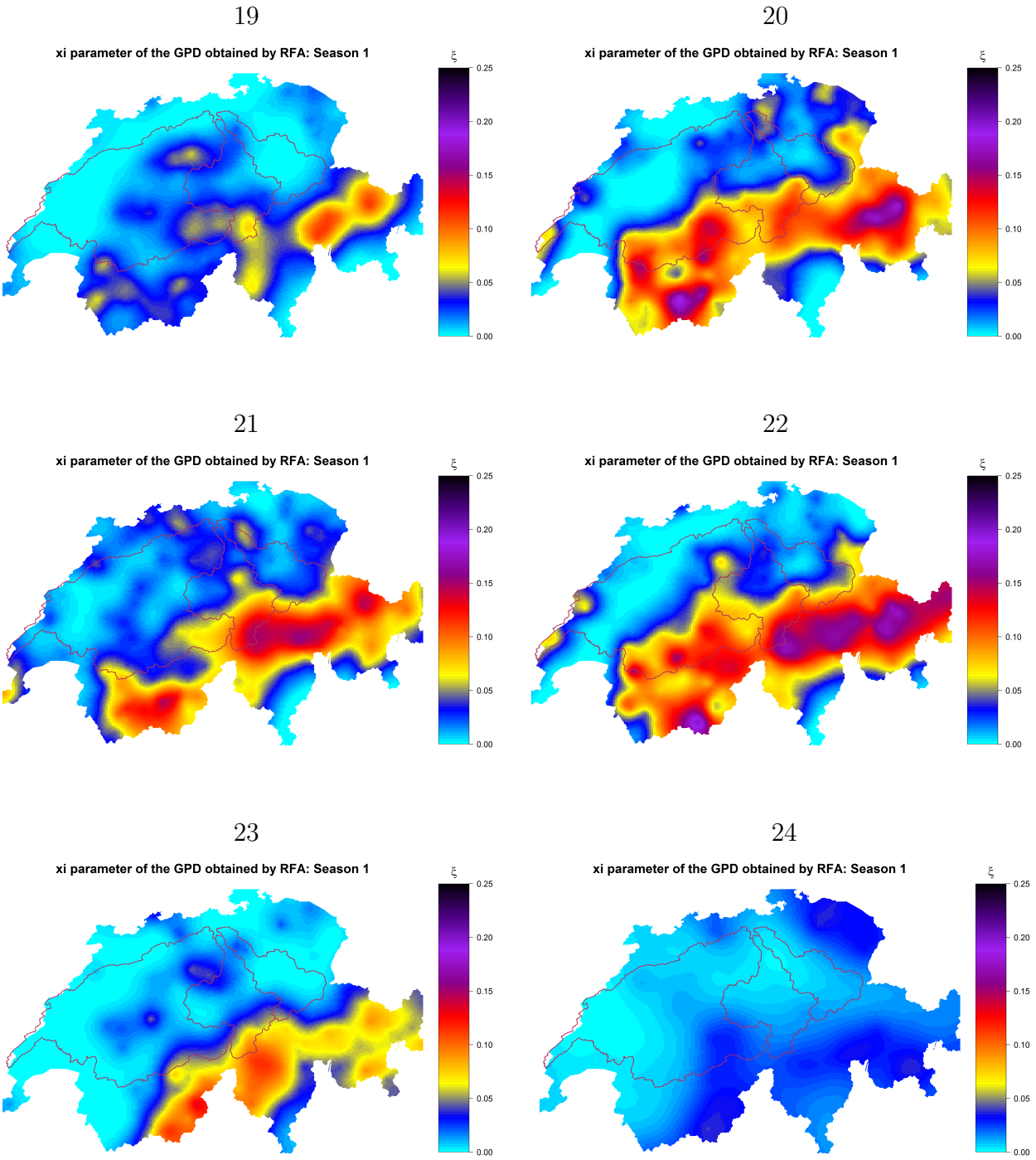


Figure 7: Maps of bootstrap ξ estimates (Winter, sample 19 to sample 24)

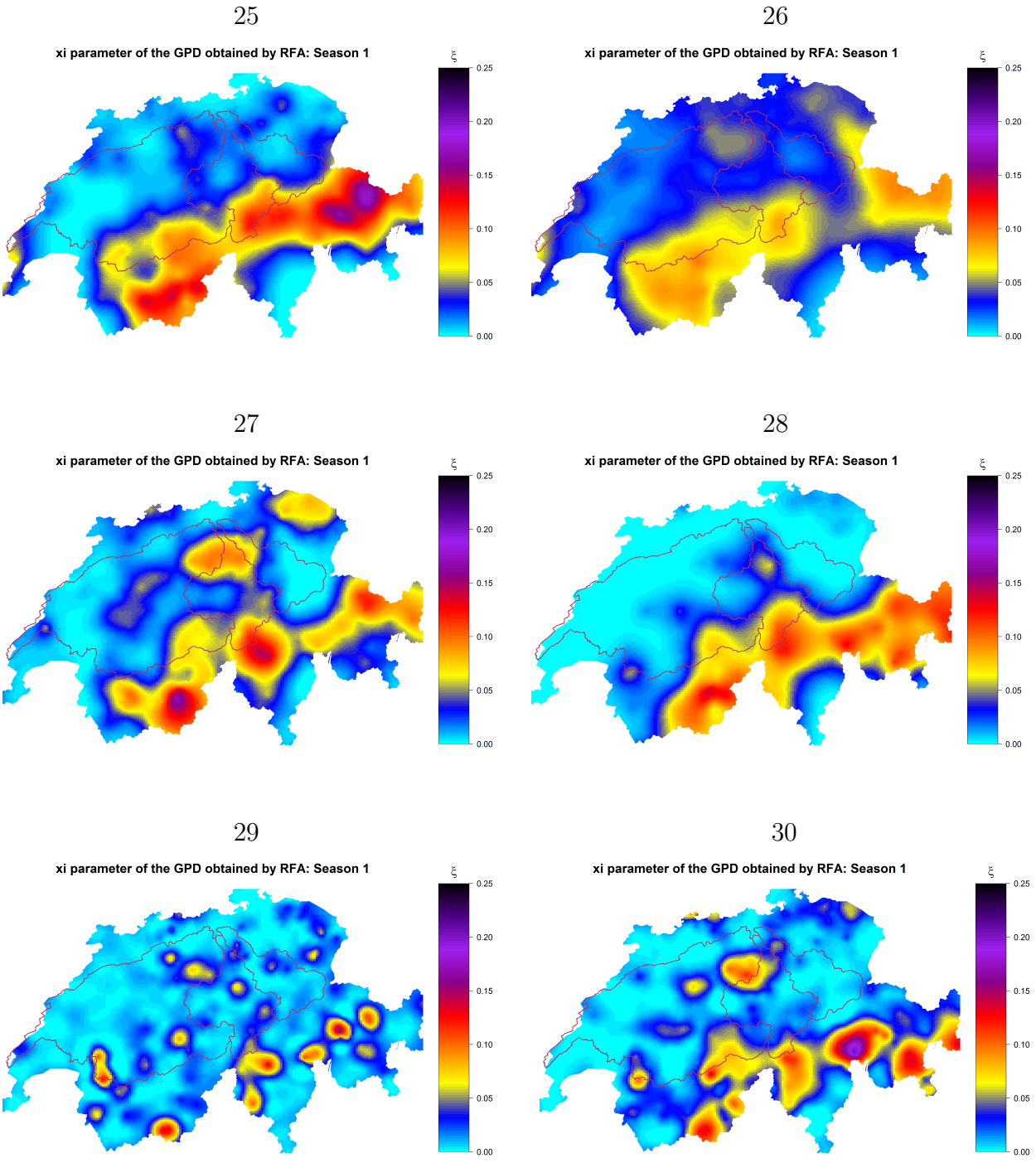


Figure 8: Maps of bootstrap ξ estimates (Winter, sample 25 to sample 30)

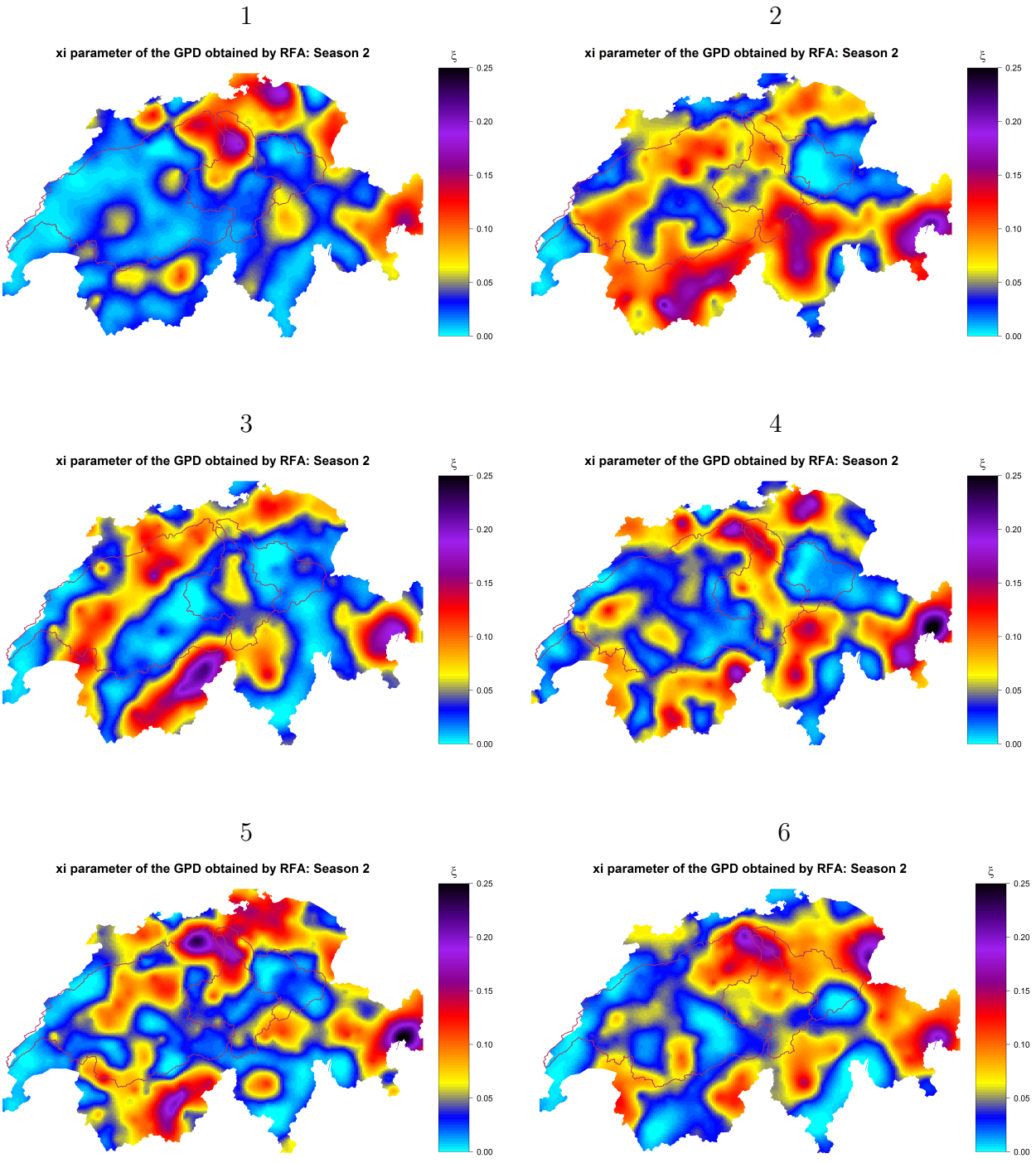
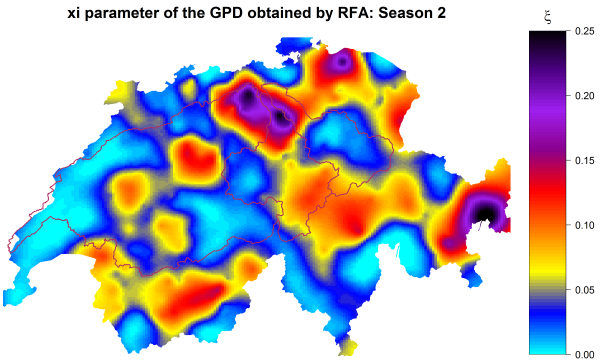
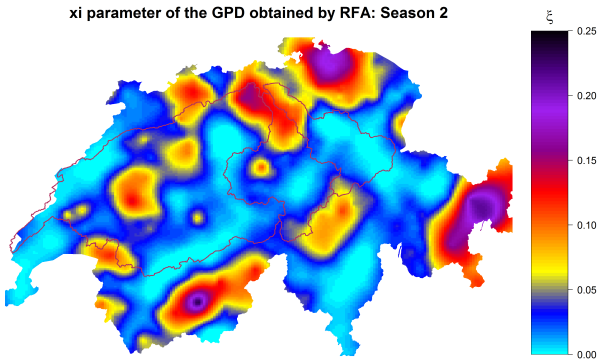


Figure 9: Maps of bootstrap ξ estimates (Spring, sample 1 to sample 6)

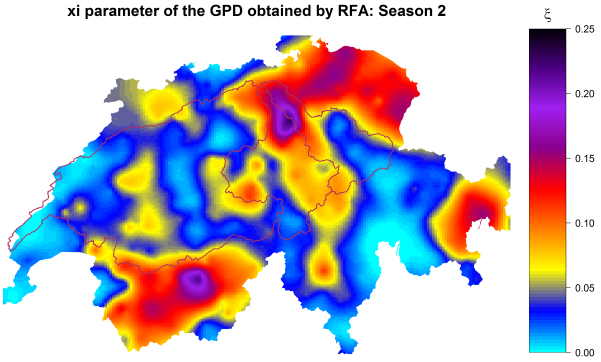
7



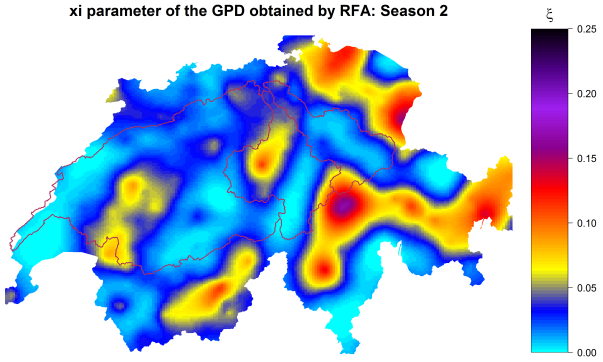
8



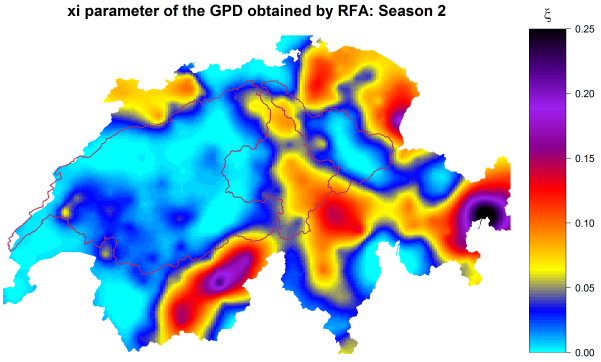
9



10



11



12

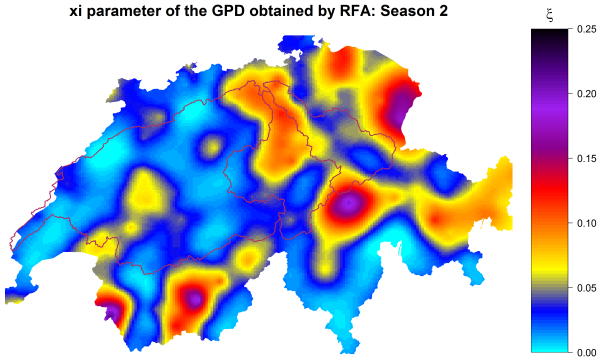
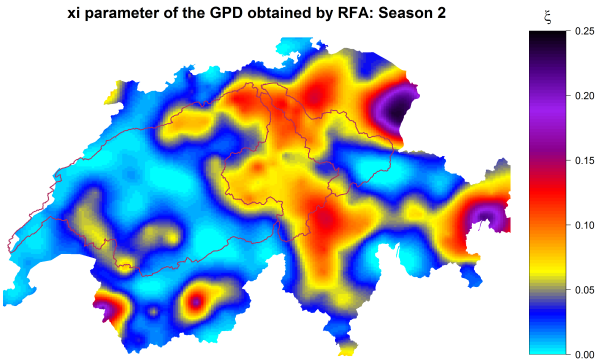
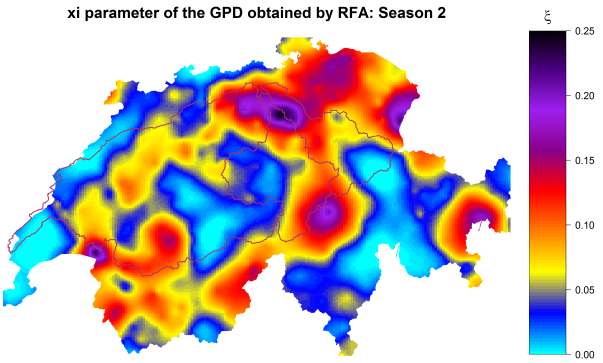


Figure 10: Maps of bootstrap ξ estimates (Spring, sample 7 to sample 12)

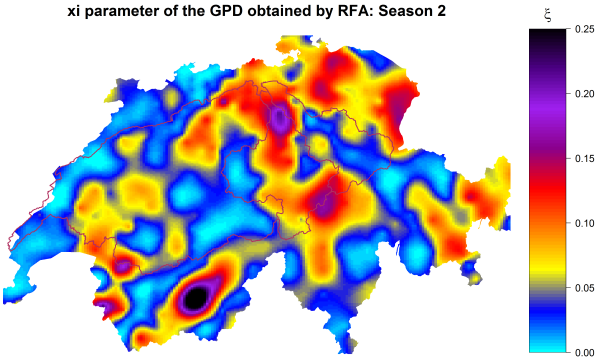
13



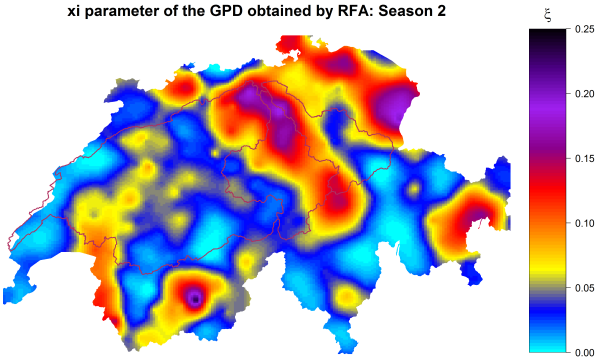
14



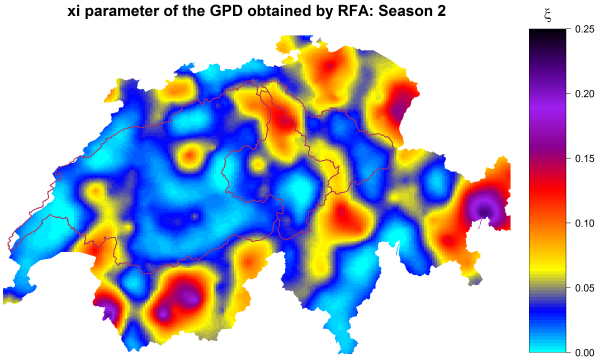
15



16



17



18

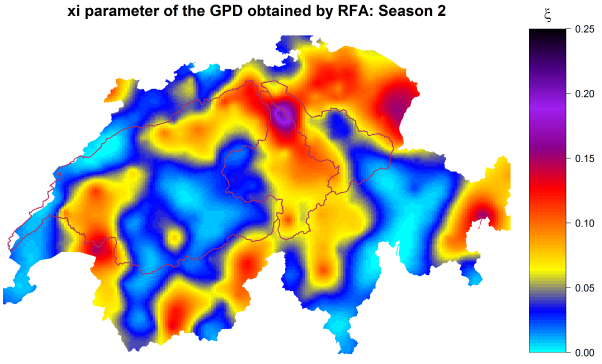
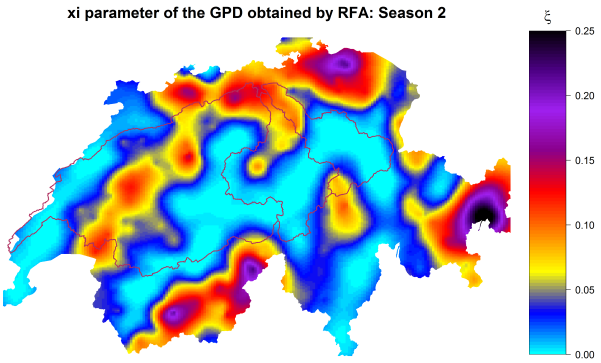
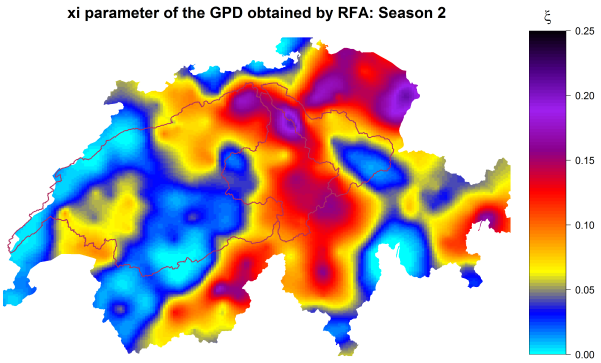


Figure 11: Maps of bootstrap ξ estimates (Spring, sample 13 to sample 18)

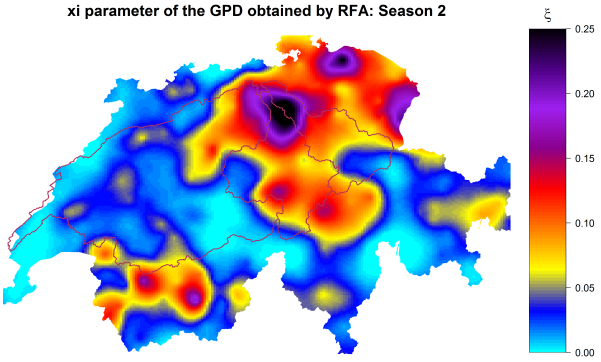
19



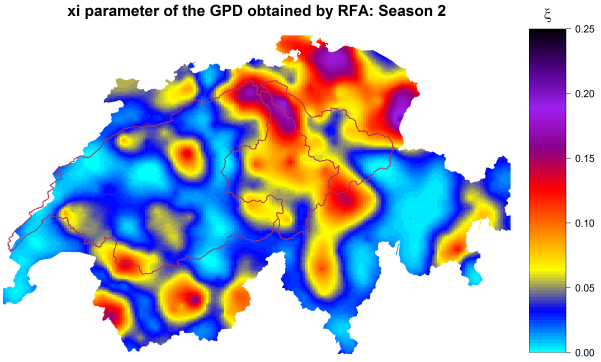
20



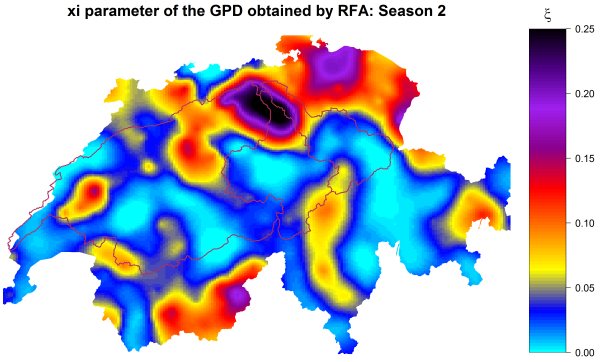
21



22



23



24

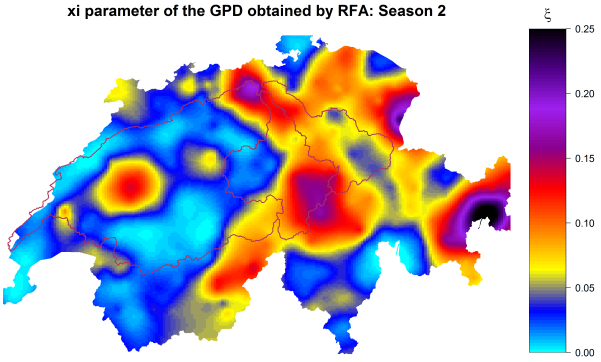


Figure 12: Maps of bootstrap ξ estimates (Spring, sample 19 to sample 24)

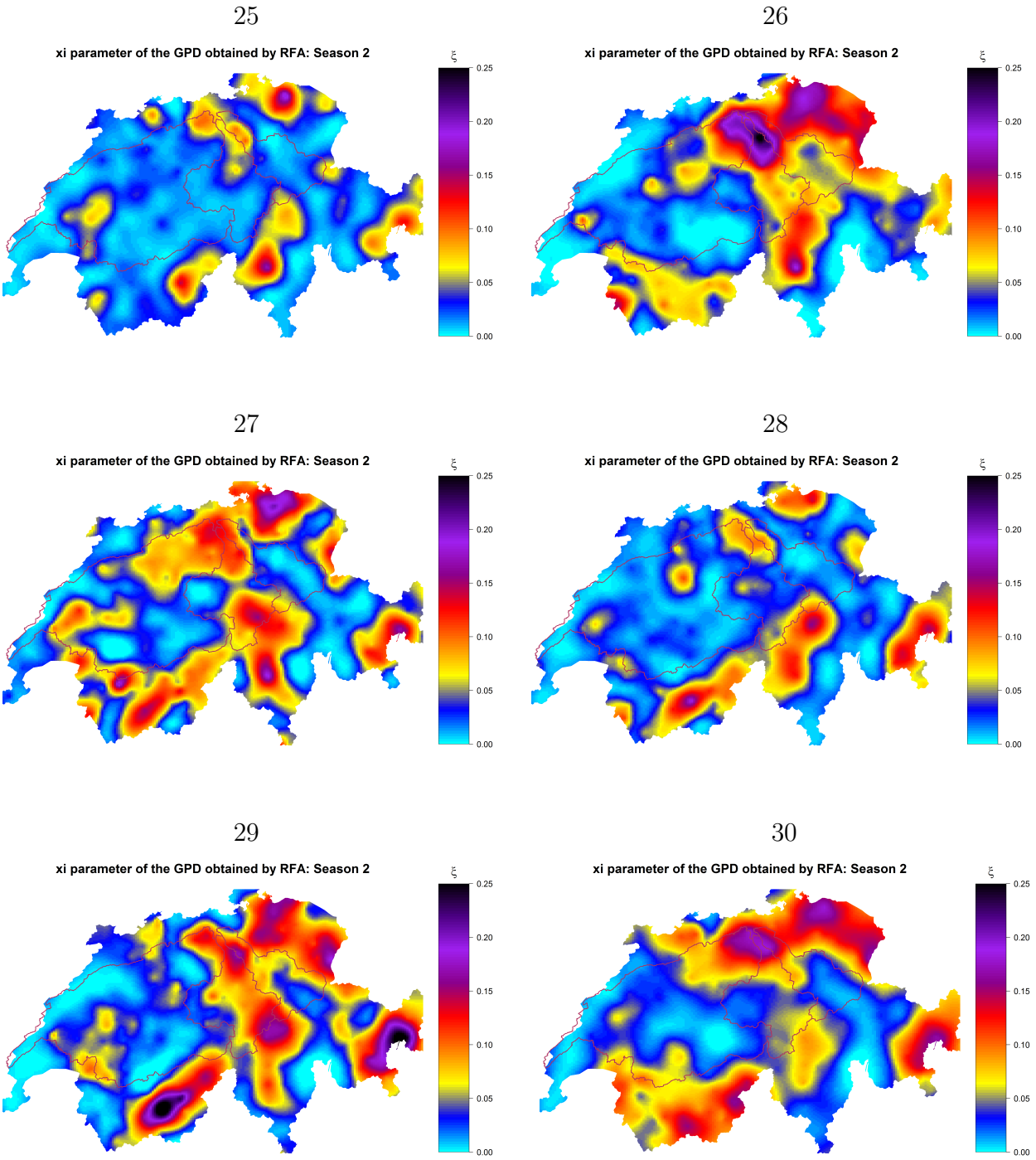


Figure 13: Maps of bootstrap ξ estimates (Spring, sample 25 to sample 30)

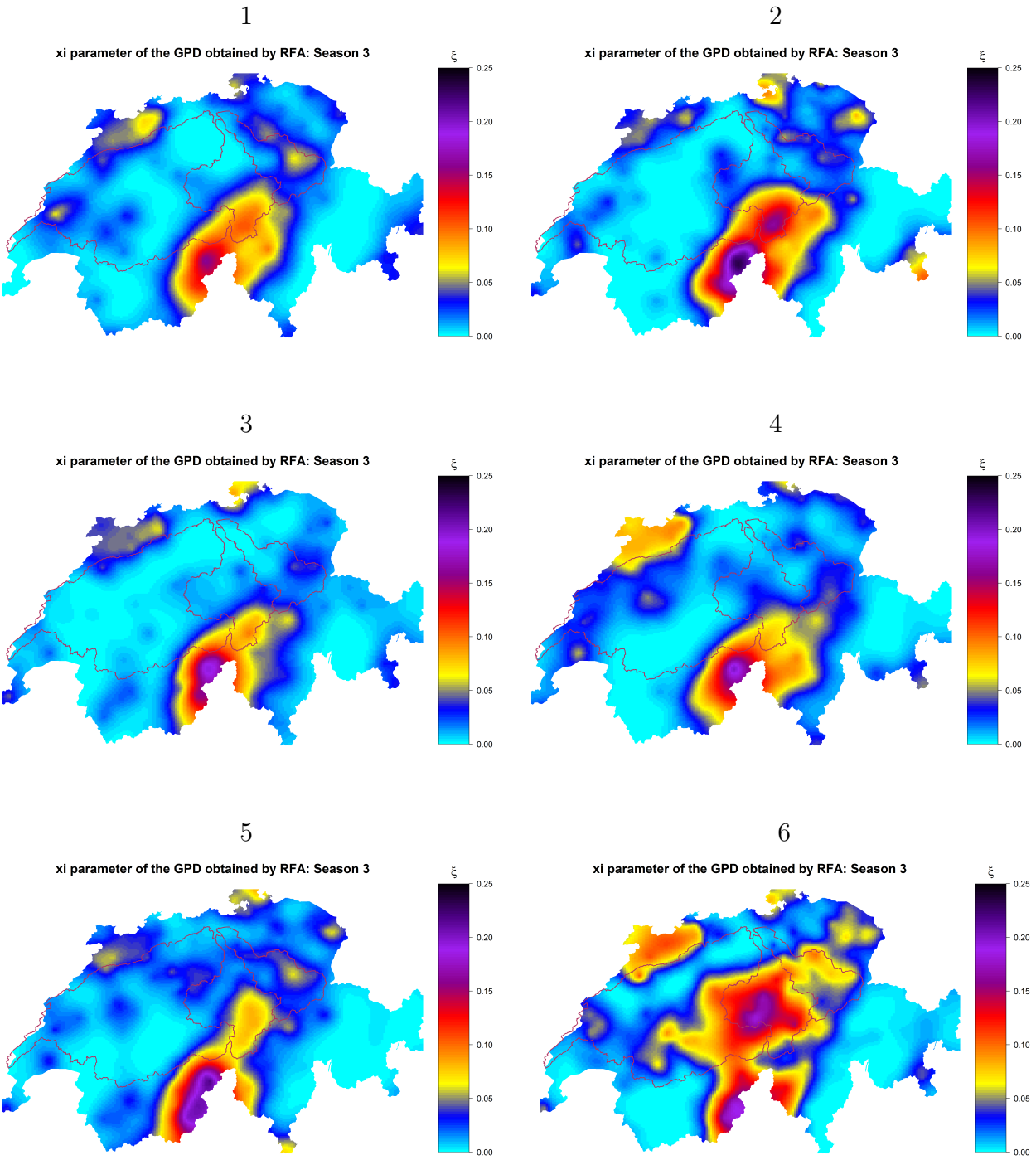


Figure 14: Maps of bootstrap ξ estimates (Summer, sample 1 to sample 6)

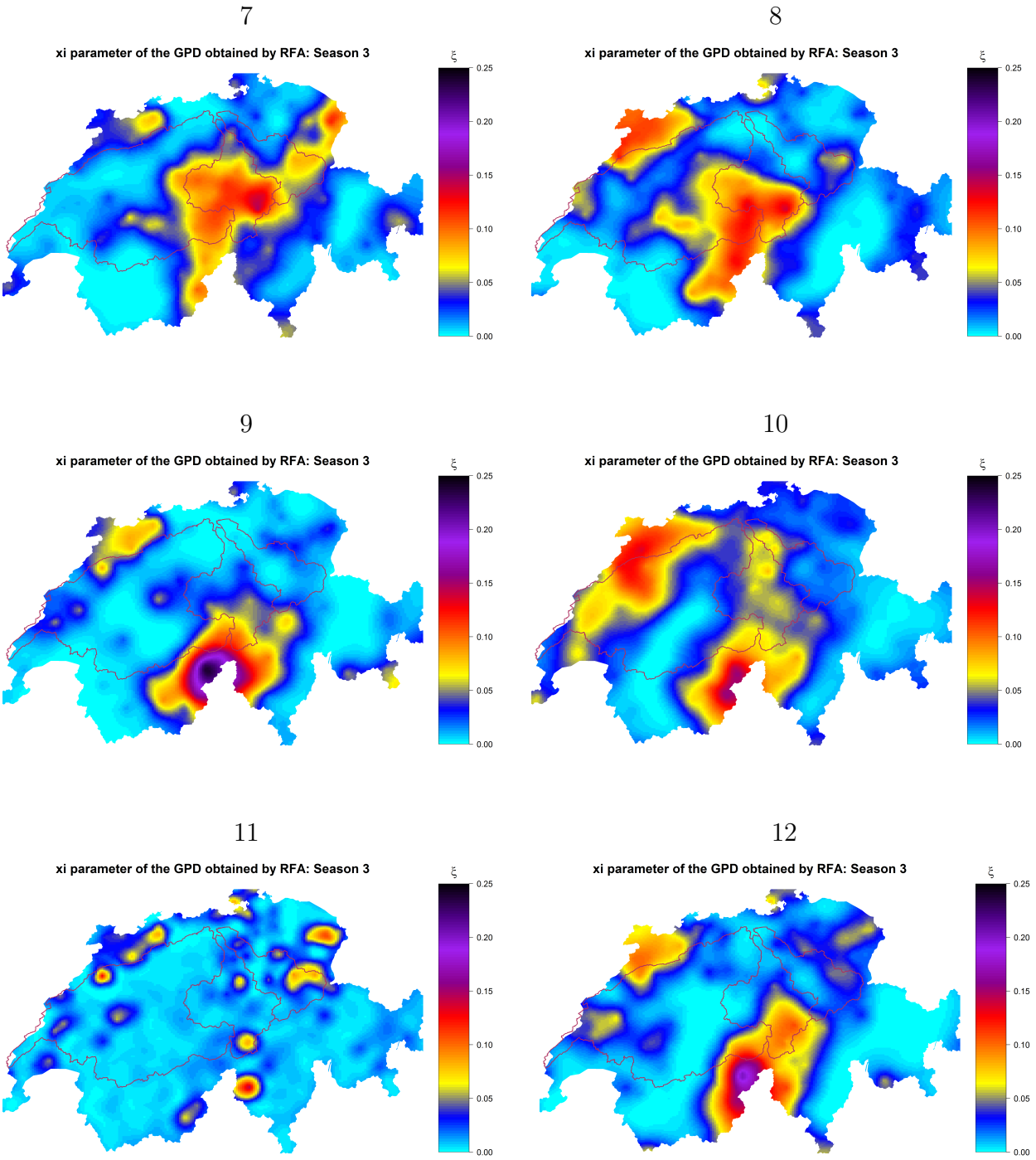
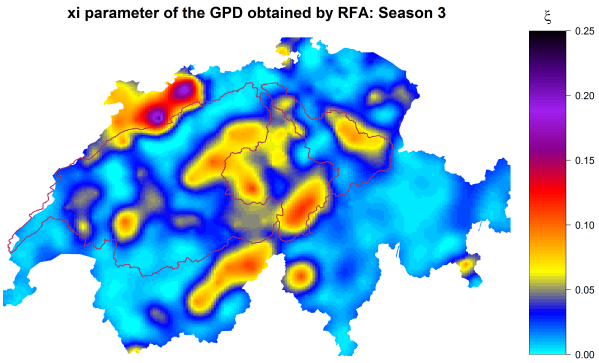
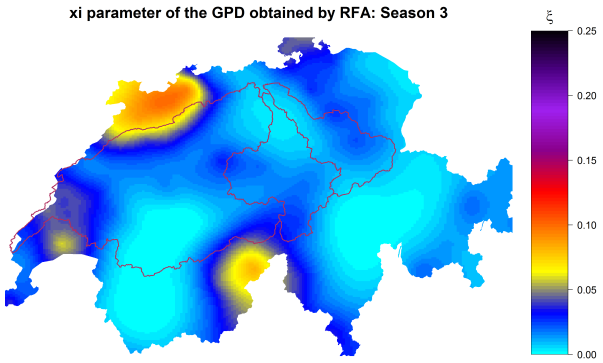


Figure 15: Maps of bootstrap ξ estimates (Summer, sample 7 to sample 12)

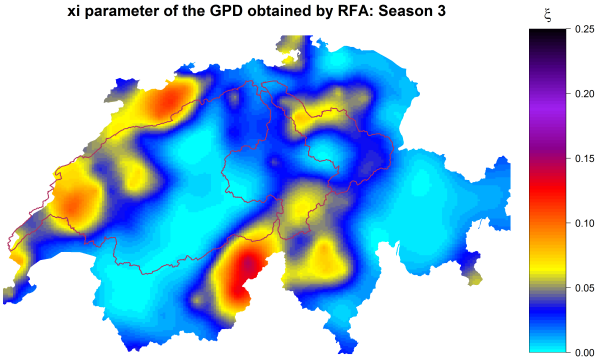
13



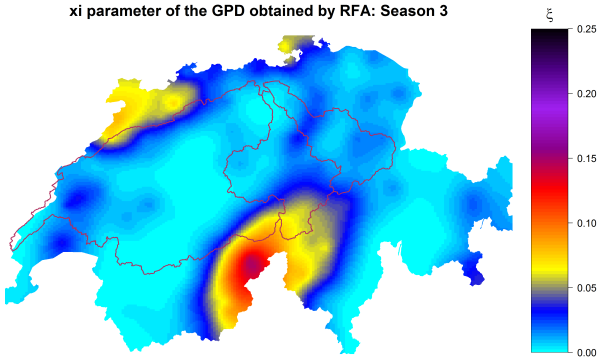
14



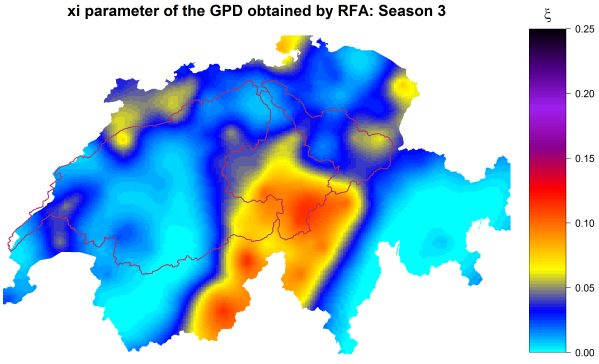
15



16



17



18

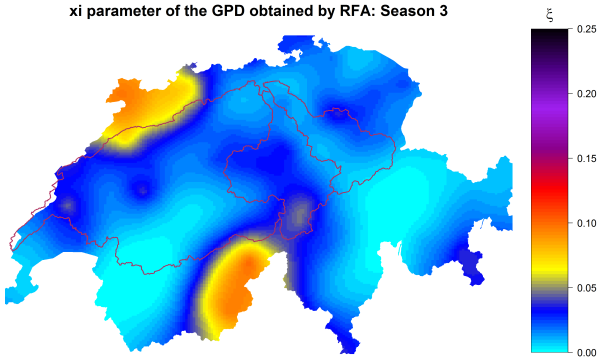


Figure 16: Maps of bootstrap ξ estimates (Summer, sample 13 to sample 18)

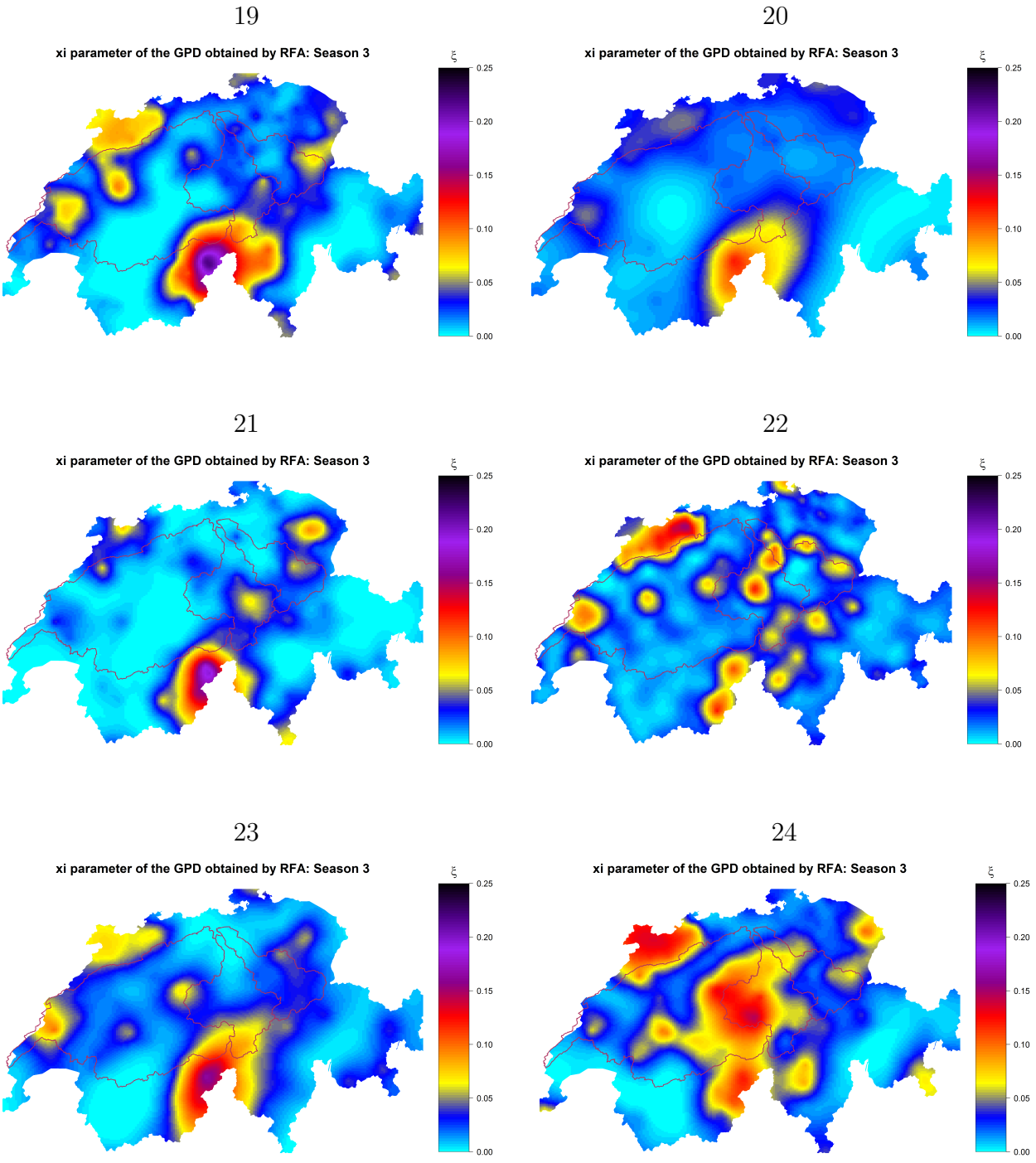


Figure 17: Maps of bootstrap ξ estimates (Summer, sample 19 to sample 24)

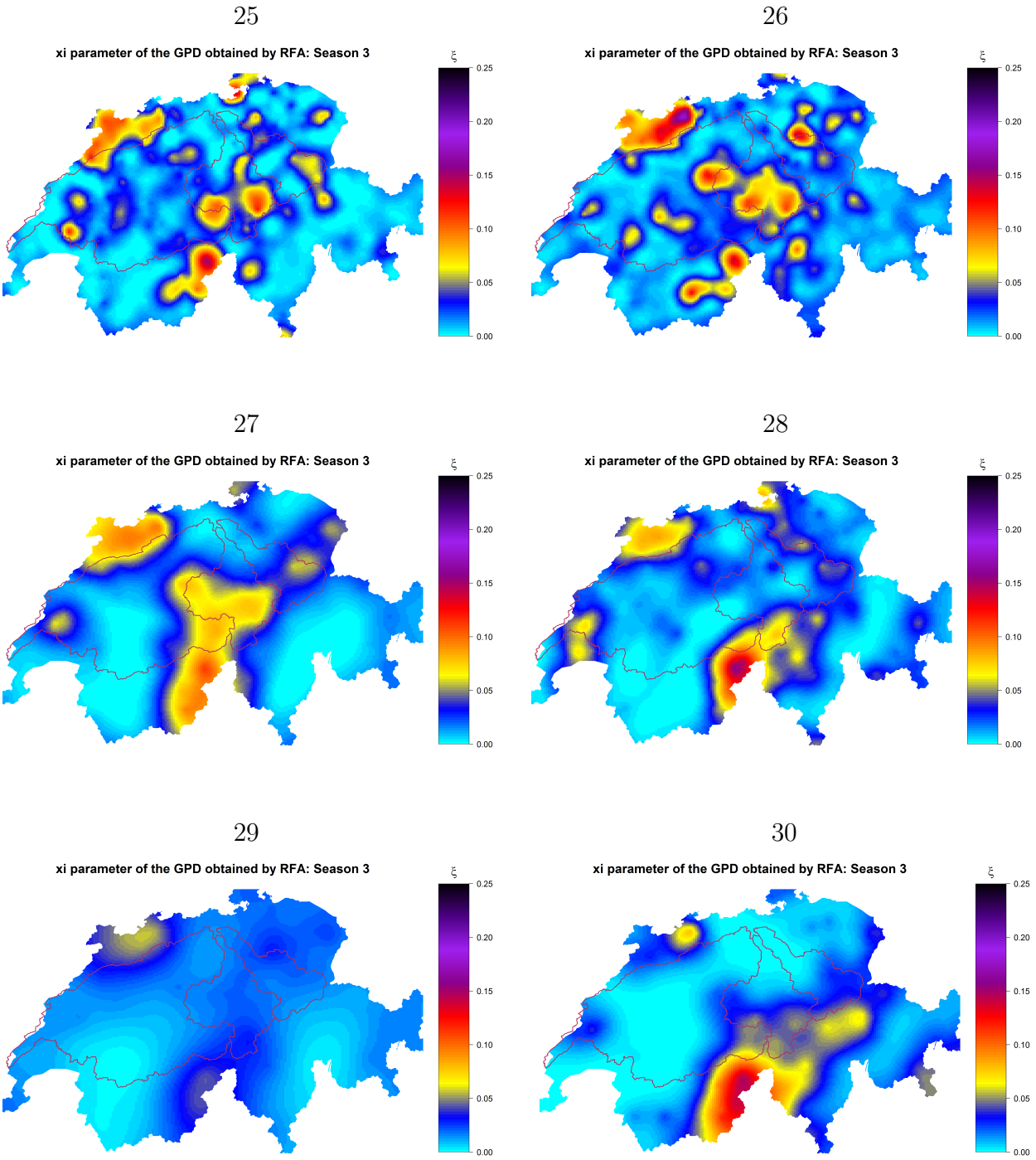


Figure 18: Maps of bootstrap ξ estimates (Summer, sample 25 to sample 30)

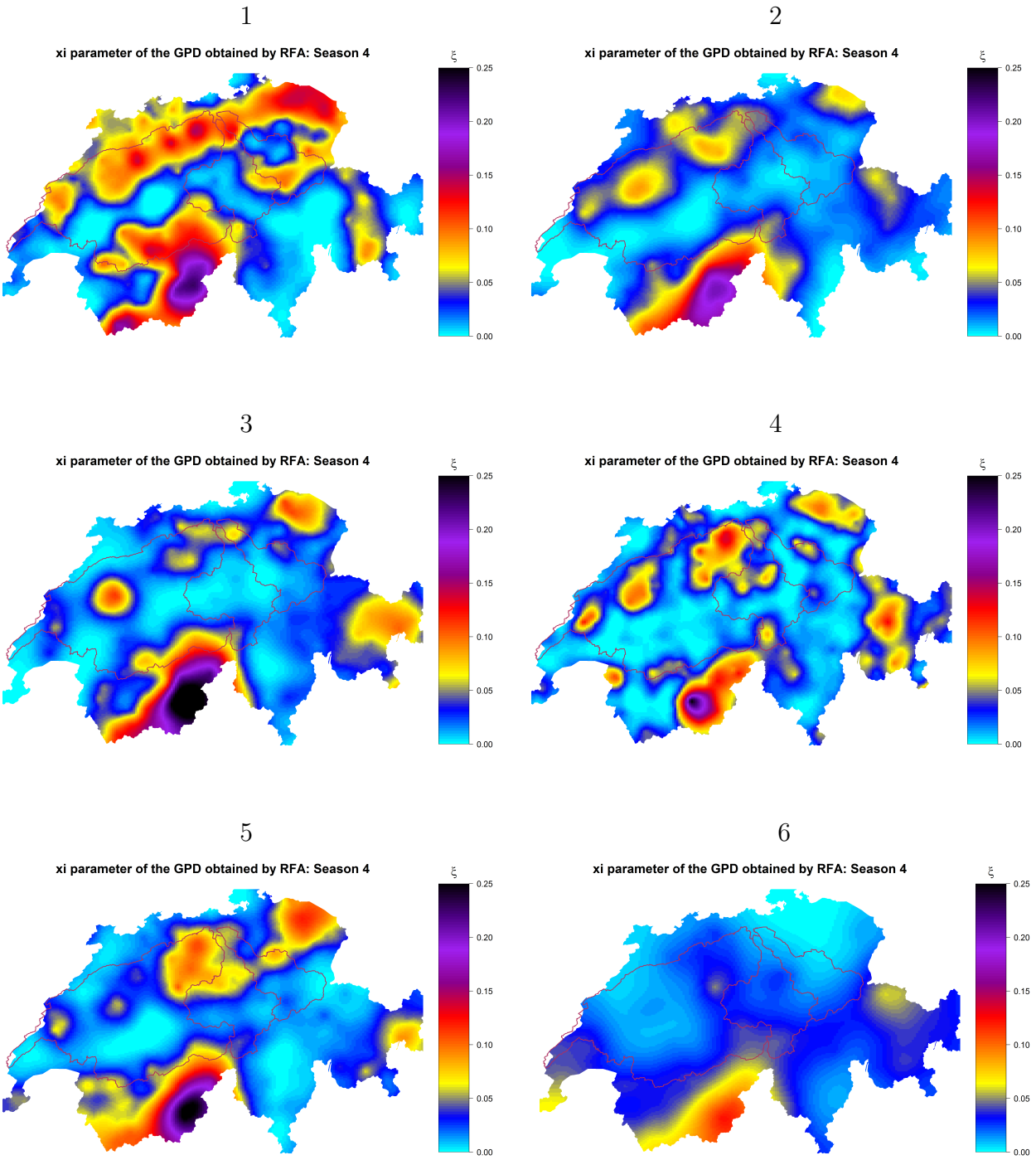
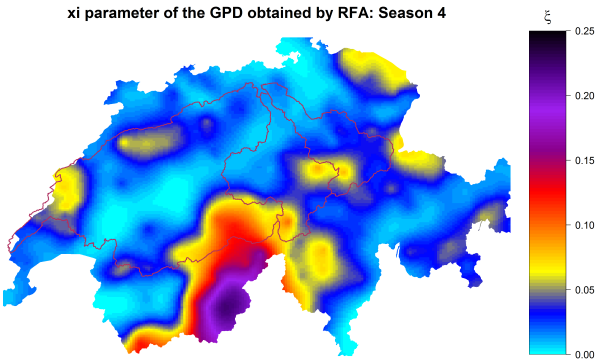
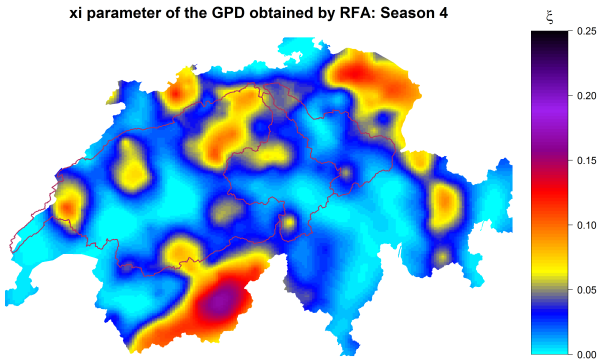


Figure 19: Maps of bootstrap ξ estimates (Autumn, sample 1 to sample 6)

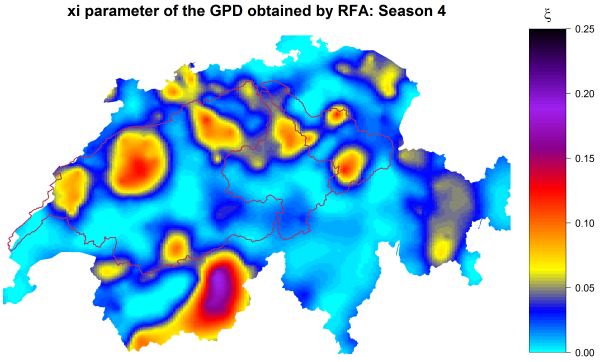
7



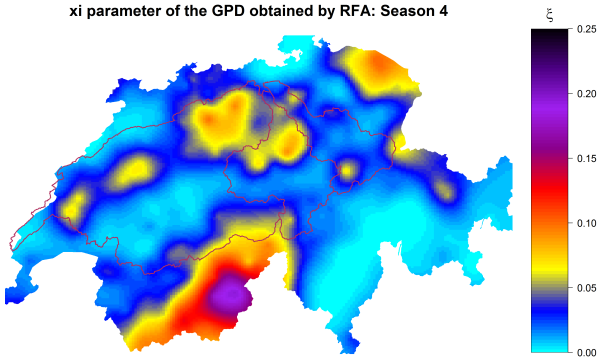
8



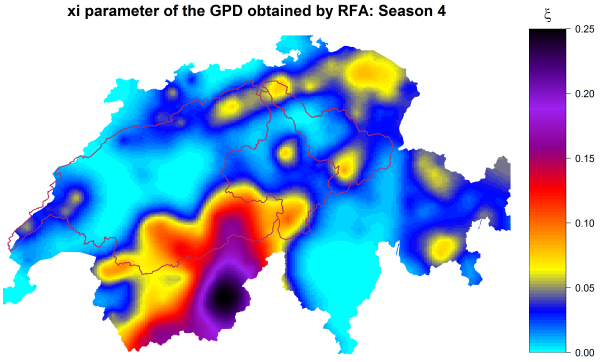
9



10



11



12

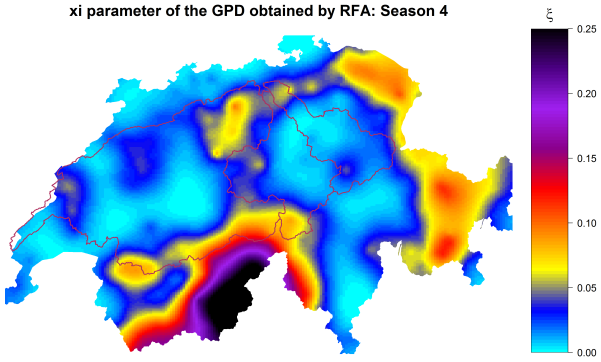


Figure 20: Maps of bootstrap ξ estimates (Autumn, sample 7 to sample 12)

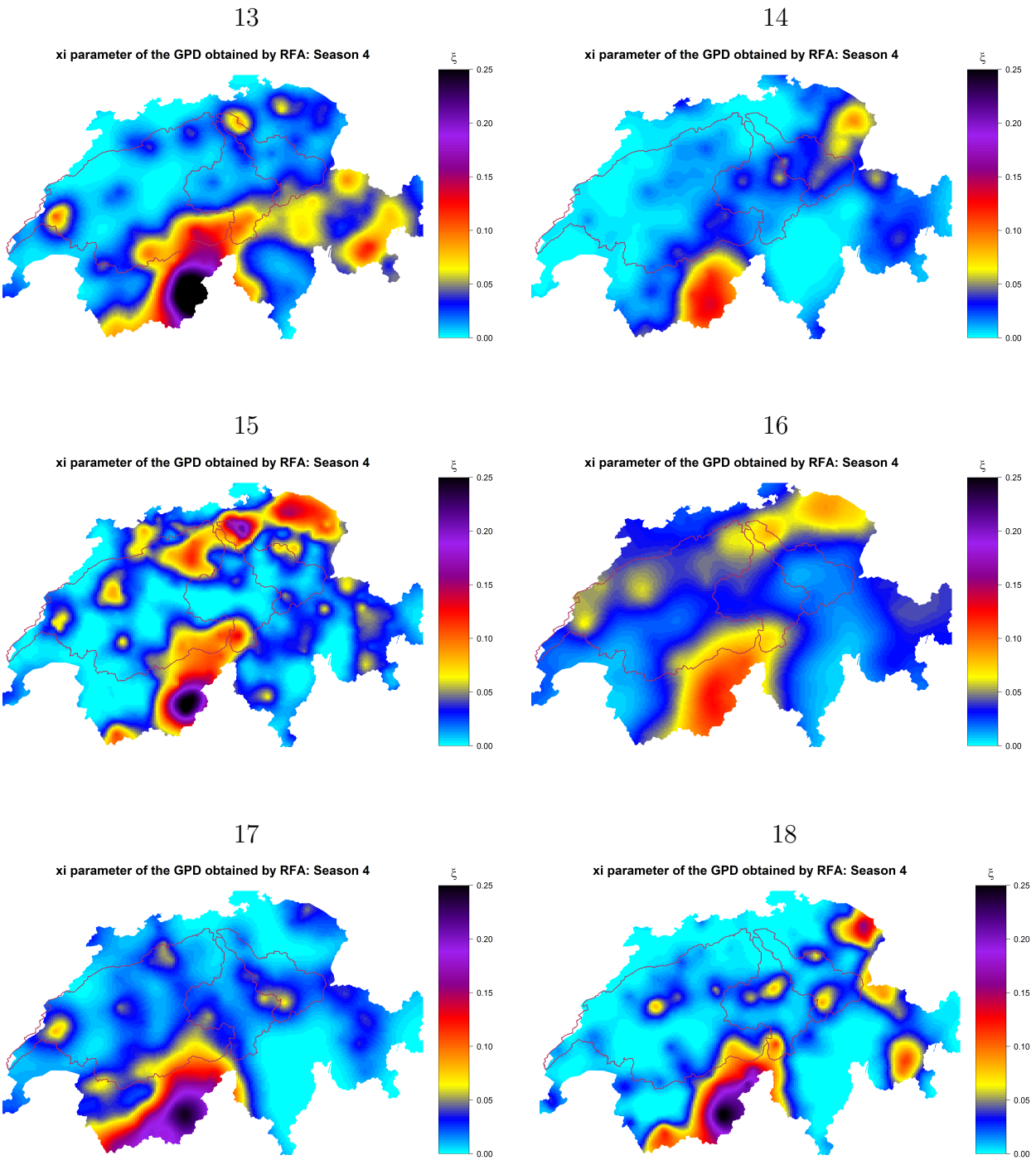
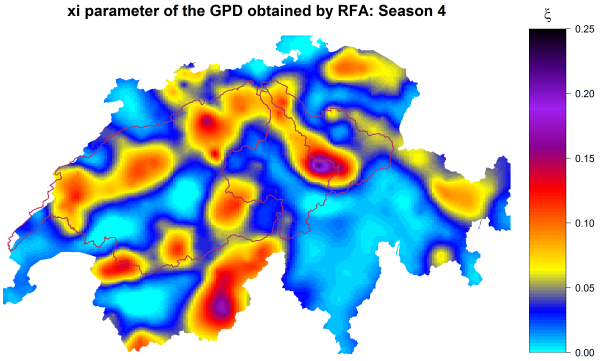
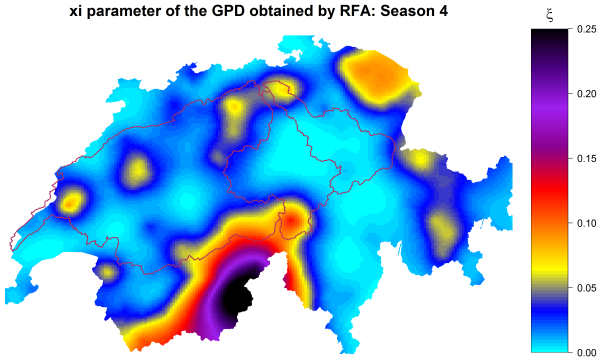


Figure 21: Maps of bootstrap ξ estimates (Autumn, sample 13 to sample 18)

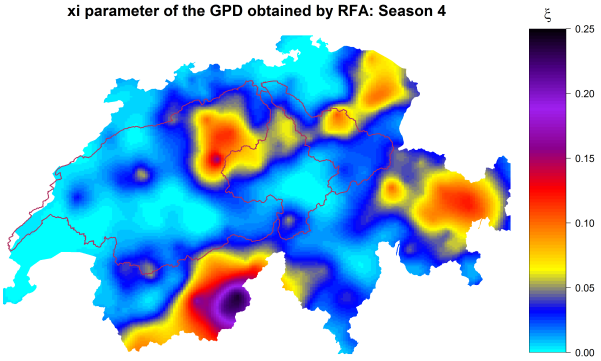
19



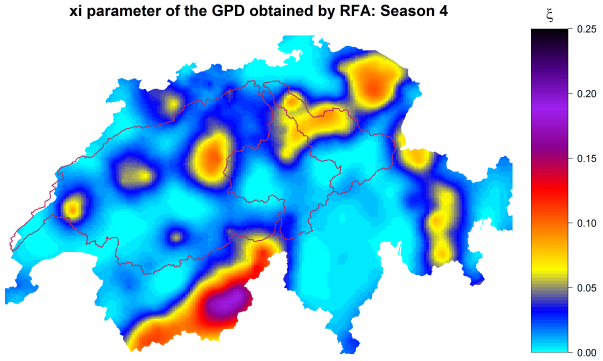
20



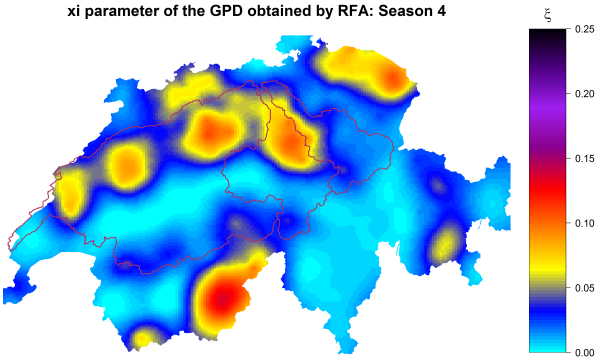
21



22



23



24

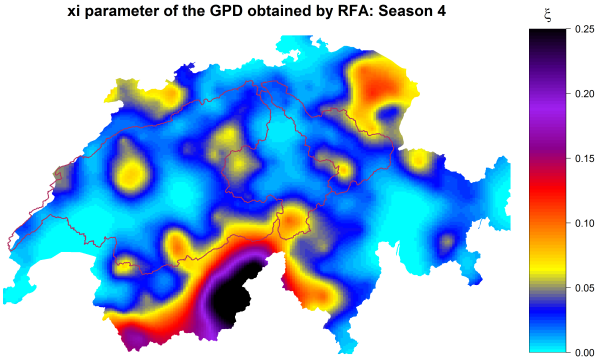
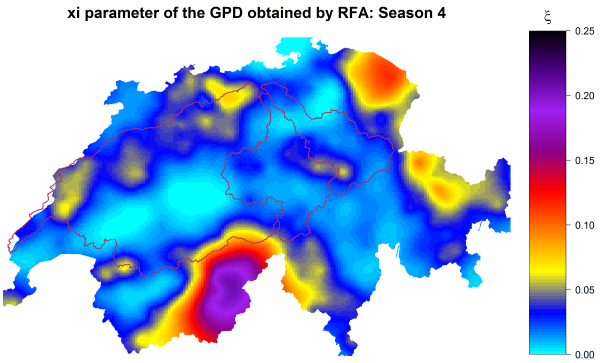
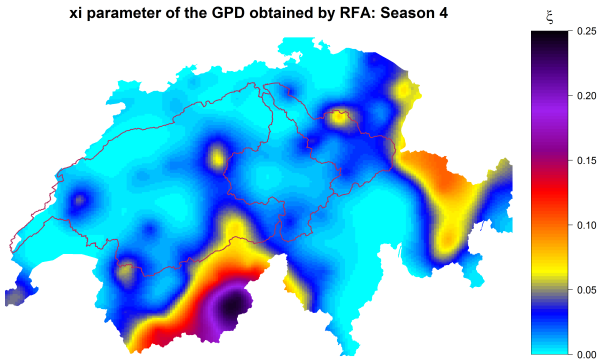


Figure 22: Maps of bootstrap ξ estimates (Autumn, sample 19 to sample 24)

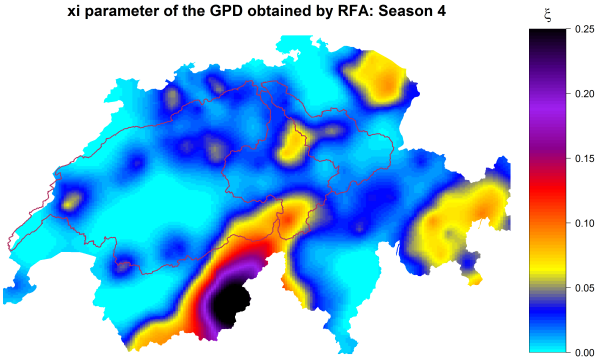
25



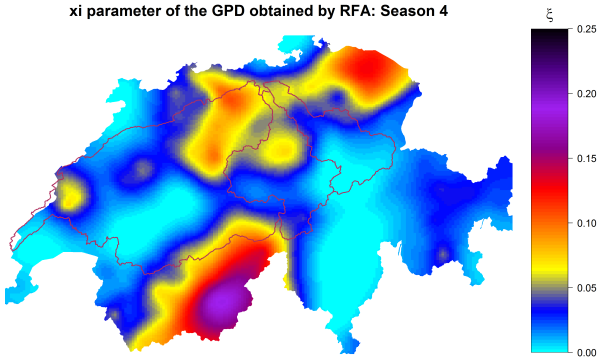
26



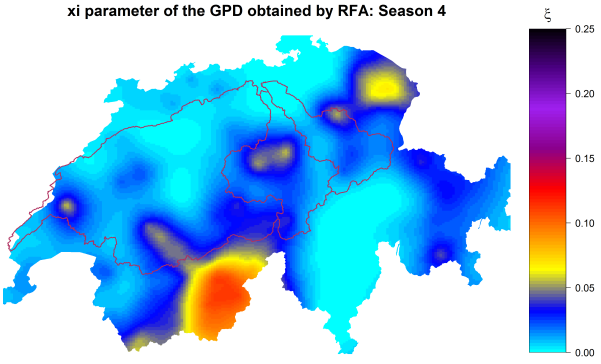
27



28



29



30

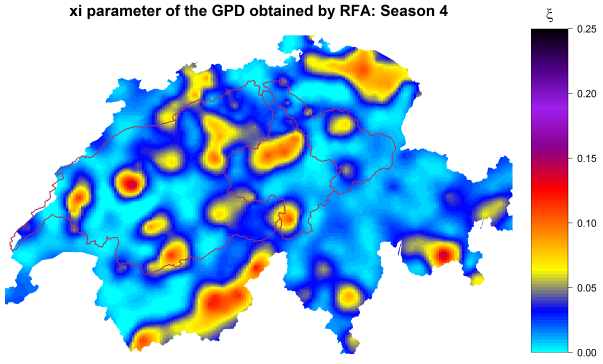


Figure 23: Maps of bootstrap ξ estimates (Autumn, sample 25 to sample 30)

10.3.6 Type of asymptotic (spatial) dependence in precipitation extremes

EXAR

Working report

July 10, 2019

Asymptotical dependence of rainfall extremes in Aare basin

Gilles Nicolet, Damien Raynaud, Guillaume Evin, Jérémy Chardon, Anne-Catherine Favre (Pr.) and Benoit Hingray (Res. Sc.)

1 Introduction

The issue of asymptotical dependence/independence of extreme precipitation is crucial when one deals with events with very large return periods. In the case of asymptotical dependence, the dependence strength does not change with the degree of extremeness. On the contrary, in the case of asymptotical independence, the dependence decreases as the events become more extreme. A false assumption of asymptotical dependence may lead to overestimated the probability of occurrence of a extreme event and, symmetrically, a incorrect hypothesis of asymptotical independence may lead to underestimated the risk.
Several studies suggest that extreme precipitations in Switzerland would be asymptotically independent. Davison et al. (2012, 2013) investigated extreme precipitation in the Swiss Plateau and Thibaud et al. (2013) applied threshold-based model to extreme precipitation in the Swiss Alps. These three articles came to the conclusion to the asymptotically independent models outperform asymptotically dependent models to represent extreme precipitation data. Nevertheless, one should note that all these studies deal with extreme precipitation during Summer.
We discuss here the issue of asymptotical dependence/independence of precipitation in Switzerland, and how it is dealt in GWEX. Firte, we define in Section 2 the asymptotical dependence/independence and presents the diagnostic tools to try to distinguish these two cases. Then, we apply these dianostic tools in Section 3 with the aim to determine the type of asympotical dependence of extreme precipitations in Switerland. Then, we draw conclusions for the GWEX model in Section 4.

2 Diagnostic tools for asymptotical dependence/independence

Coles et al. (1999) proposes diagnostic tools to determine if a dataset should be consider as asymptotically independent or not. These diagnostic tools are not easy to use, and to identify asymptotic dependence/independence is generally a tricky problem due to the length of the samples at our disposal comparing to the asymptotic nature of this issue.
Let (U, V) be two uniform random variables (eventually, transformed into uniform variables if necessary). A classical measure of the extremal dependence between U and V is

$$\chi = \lim_{u \to 1} \mathbb{P}\left(V > u | U > u\right). \tag{1}$$

which is the probability of one variable to be extreme when the other is extreme. If $\chi = 0$, U and V are asymptotically independent.
We define

$$\chi(u) = 2 - \frac{\log \mathbb{P}\left(U < u, V < u\right)}{\log \mathbb{P}\left(U < u\right)} \text{ for } 0 \leq u \leq 1. \tag{2}$$

Asymptotically,

$$\chi(u) \sim \mathbb{P}\left(V > u | U > u\right) \tag{3}$$

and therefore we have

$$\chi = \lim_{u \to 1} \chi(u). \tag{4}$$

The function $\chi(u)$ can be interpreted as a measure of dependence dependent of the threshold u. The sign of $\chi(u)$ indicates whether U and V are positively or negatively associated. Thus, $\chi(u) = 0$ signifies independent variables and $\chi(u) = 1$ means perfect dependence.
Coles et al. (1999) recommends to use the measure $\chi(u)$ together with a second measure noted $\bar{\chi}(u)$ and defined by

$$\bar{\chi}(u) = \frac{2 \log \mathbb{P}(U > u)}{\log \mathbb{P}(U > u, V > u)} - 1 \text{ for } 0 \leq u \leq 1. \tag{5}$$

We have $-1 < \bar{\chi}(u) \leq 1$ for all $0 \leq u \leq 1$. We also define

$$\bar{\chi} = \lim_{u \to 1} \bar{\chi}(u). \tag{6}$$

$\bar{\chi} = 1$ signifies asymptotically dependent variables. In the case of asymptotically independent variables ($\bar{\chi} < 1$), $\bar{\chi}$ measure the strength of the dependence.
Coles et al. (1999) suggests to use the two measures $(\chi, \bar{\chi})$ together for asymptotical dependence/independence diagnostics: $(\chi > 0, \bar{\chi} = 1)$ means asymptotic dependence while $(\chi = 0, \bar{\chi} < 1)$ means asymptotic independence.

3 Asymptotical dependence/independence of extreme precipitation in Switzerland

As we consider 105 stations in the Aare basin and that the GWEX model is apply for each month, we have to consider $\frac{105 \times 104}{2} = 5460$ pairs of stations for each month. Figure 1 shows that one may observes very different cases amoung these 5460 pairs of stations:

- pairs of stations with weak or no tail dependence (e.g., GOS and KUE);
- pairs of stations with strong tail dependence but this tail dependence decreases with extremeness (asymptotically independent) (e.g., ALT and ENT);
- pairs of stations with strong tail dependence and asymptotically dependent (e.g., AIE and BER).

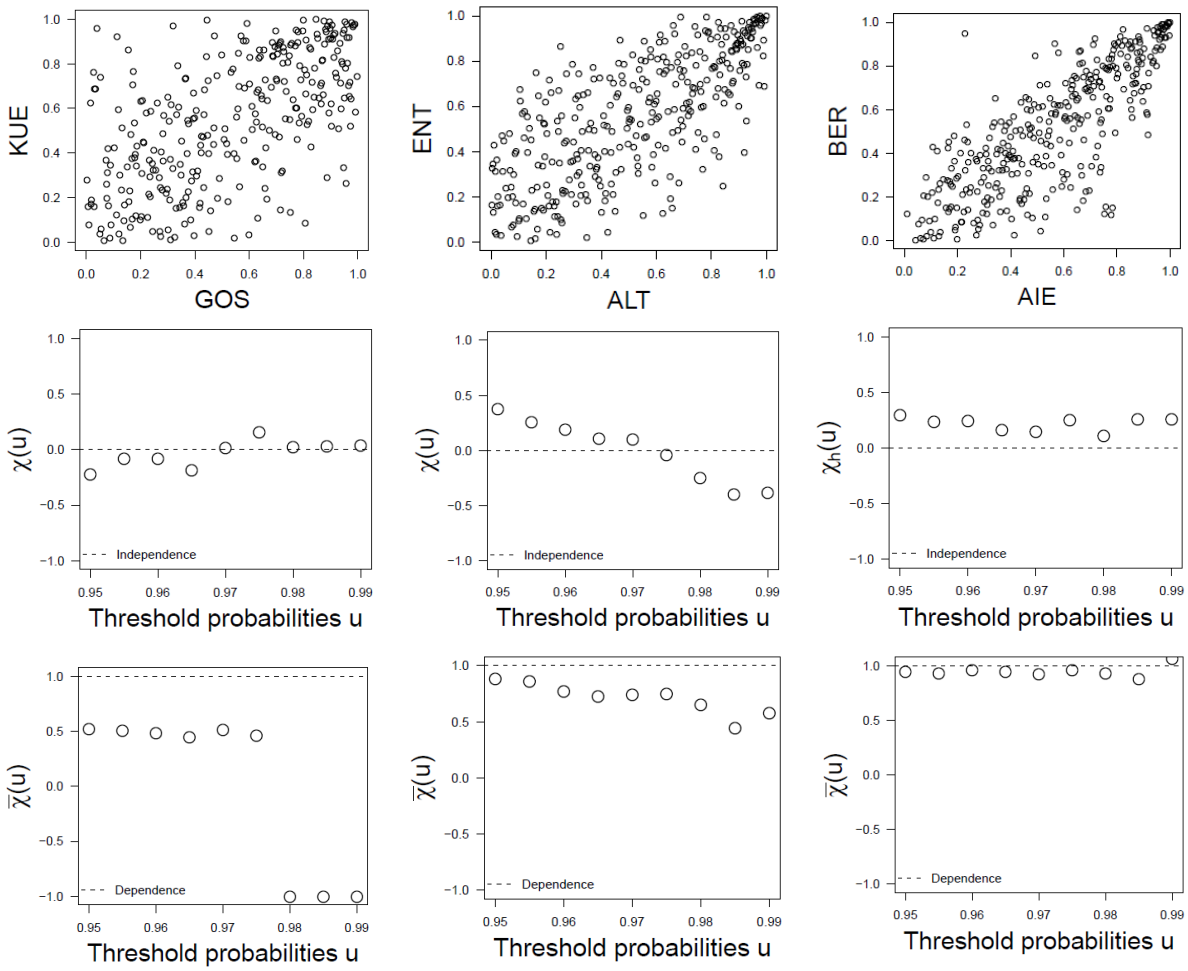


Figure 1: Scatterplots (up), $\chi(u)$ measure (middle) and $\bar{\chi}(u)$ measure (bottom) for three pairs of stations: Göchenen (GOS) and Küsnacht (KUE) (left), Altdorf (ALT) and Entlebuch (ENT) (middle), and Affoltern im Emmental (AIE) and Bern/Zollikofen(BER) (right).

In order to have an overall view of asymptotical dependence/independence, we arbitrarily choose a criteria ($\chi(0.99) > 0.1$ and $\bar{\chi}(0.99) > 0.9$) to determine asymptotical dependence. Figures 2 and 3 show the asymptotically dependent pairs of stations according this criteria. We observe a strong seasonality in asymptotic dependence with many asymptotical dependent relations from November to March, and with a lesser degree, in May, in August (in the Plateau) and in October (in the Alps). Extreme precipitations seem to be asymptotically independent during the summer, which is consistent with Davison et al. (2012, 2013); Thibaud et al. (2013).

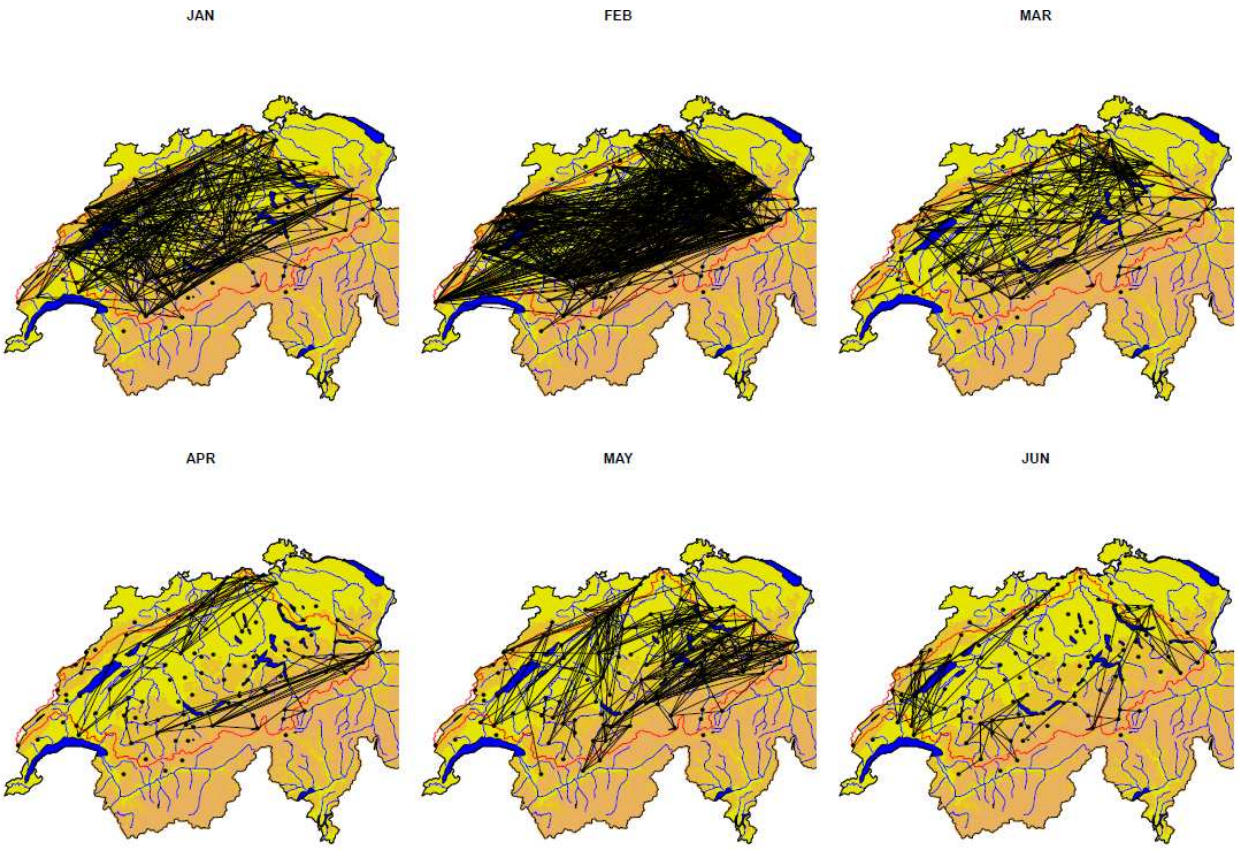


Figure 2: Asymptotically dependent pairs of stations (from January to June). Two stations are linked if we have $\chi(0.99) > 0.1$ and $\bar{\chi}(0.99) > 0.9$.

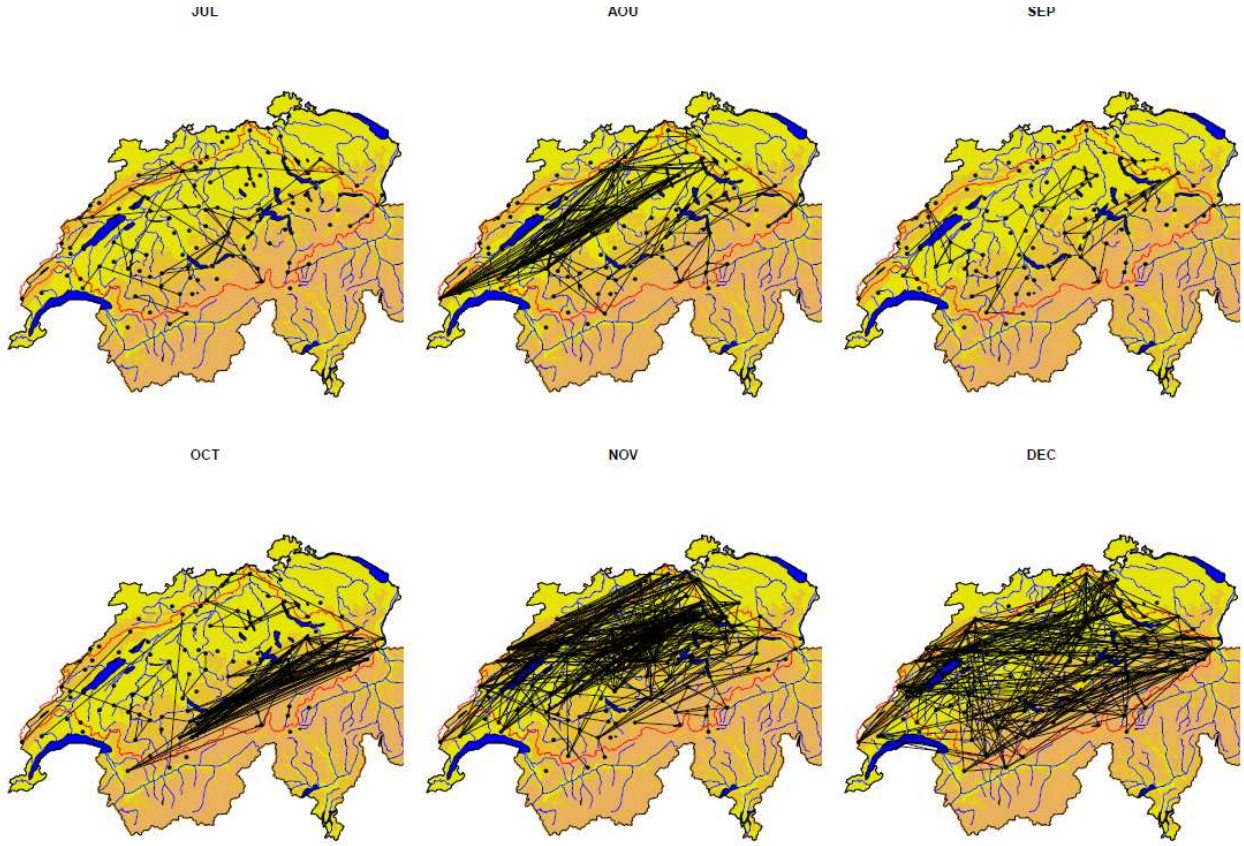


Figure 3: Asymptotically dependent pairs of stations (from July to December). Two stations are linked if we have $\chi(0.99) > 0.1$ and $\bar{\chi}(0.99) > 0.9$.

4 Asymptotical dependence/independence in GWEX model

The GWEX model (Evin et al., 2018) decomposes the precipitation process $P_t(k)$ into two independent processes: an occurrence process $X_t(k)$ and an intensity process $Y_t(k)$

$$P_t(k) = Y_t(k)X_t(k) \tag{7}$$

for each station k at each day t. A extended generalized Pareto distribution (Papastathopoulos and Tawn, 2013) F_Y is fitted to $Y_t(k)$ at each station k. Then, the gaussianized process $\mathbf{Z}_t$ defined by

$$Z_t(k) = \Phi^{-1}\left[F_Y\left\{Y_t(k)\right\}\right] \tag{8}$$

with Φ the cumulative distribution function of the standard normal distribution, is modeled through a MAR(1) process

$$\mathbf{Z}_t = \mathbf{A}\mathbf{Z}_{t-1} + \epsilon_t \tag{9}$$

with ϵ_t an innovation term modeled by a copula. The innovation term ϵ_t is the one which control the spatial dependence of precipitation intensity in GWEX. We have investigated three copula for GWEX:

- the Gaussian copula for which extremes are independent
- the Student copula which models tail dependence between stations
- the Fisher copula (Favre et al., 2018) which models tail dependence between stations and allows asymmetry between lower tail and upper tail.

We have selected the Student copula which is able to model spatial dependence between extremes with asymptotical dependence (the Fisher copula gives very similar results). The use of the Gaussian copula is not suitable because it is necessary to model spatial dependence in extreme precipitation to avoid underestimation of spatially large extreme events. To the best of our knowledge, we do not have at our disposal in the literature copulas able to modeling dependence of extremes with asymptotical independence.
According to the two maps in Figures 2 and 3, except for specific months like June, July and September, extreme precipitations in the Aare basin seem to be asymptotical independence. Thus, the use of Student copula (or Fisher copulas) is justified from a risk mitigation point of view.

References

Coles, S., Heffernan, J., and Tawn, J. (1999). Dependence measures for extreme value analyses. *Extremes*, 2(4):339–365.

Davison, A. C., Huser, R., and Thibaud, E. (2013). Geostatistics of dependent and asymptotically independent extremes. *Mathematical Geosciences*, 45(5):511–529.

Davison, A. C., Padoan, S. A., Ribatet, M., et al. (2012). Statistical modeling of spatial extremes. *Statistical science*, 27(2):161–186.

Evin, G., Favre, A.-C., and Hingray, B. (2018). Stochastic generation of multi-site daily precipitation focusing on extreme events. *Hydrology and Earth System Sciences*, 22(1):655–672.

Favre, A.-C., Quessy, J.-F., and Toupin, M.-H. (2018). The new family of fisher copulas to model upper tail dependence and radial asymmetry: Properties and application to high-dimensional rainfall data. *Environmetrics*, 29(3):e2494.

Papastathopoulos, I. and Tawn, J. A. (2013). Extended generalised Pareto models for tail estimation. *Journal of Statistical Planning and Inference*, 143(1):131–143.

Thibaud, E., Mutzner, R., and Davison, A. C. (2013). Threshold modeling of extreme spatial rainfall. *Water resources research*, 49(8):4633–4644.

10.4 SCAMP weather generator

10.4.1 Generating new atmospheric trajectories

One of the limitations of the analogue method is the length of the synoptic weather information that is used to generate time series of local predictors. Without any additional adaption of the method, the length of the weather time series generated with SCAMP would be limited to 110 years (length of the ERA20C reanalysis). We propose here a method that extends the length of the large scale information by rearranging the synoptic weather sequences and creating new atmospheric trajectories. This method has first been proposed by (Yiou et al. 2014) to build a weather generator able to reproduce relevant characteristics for winter and summer temperature. The method we describe here relies on an analogy criterion. The only predictor involved is the geopotential height field at 1000 hPa. The analogy window is presented in Figure A. 7and corresponds to the domain used in Philipp et al., (2010) for the identification of Swiss weather time.

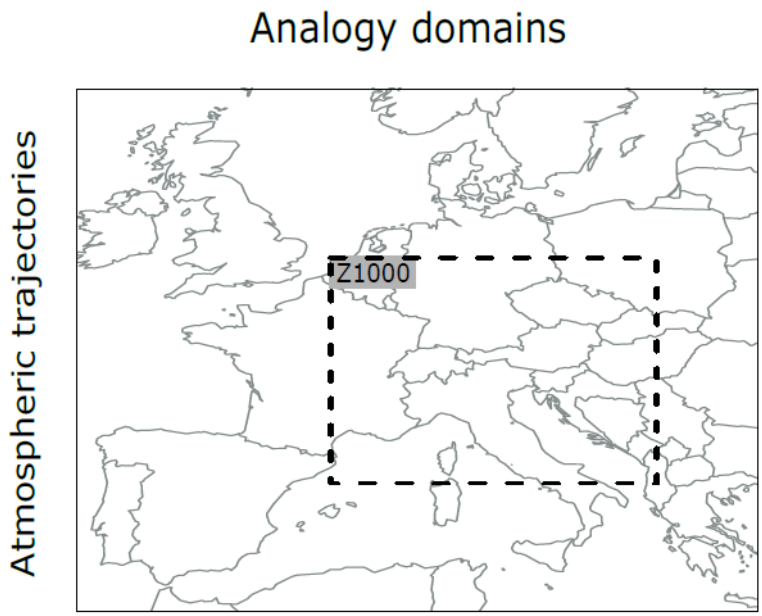


Figure A. 7 Position and dimension of the analogy window used for the construction of atmospherics trajectories. Z1000 Geopotential at 1000hPa.

The main hypothesis of atmospheric trajectories is that if two days are close analogues and if the weather patterns are heading in the same direction, then their "future" is equi-probable and one can jump for one trajectory to the other. This concept is presented on Figure A. 9 Decision tree and the different tests taking part in the construction of a new atmospheric trajectory. Day D-1 is already defined and we here explain the different option for the choice of day D. "An" stands for the analogue days.. The first line presents an observed atmospherics trajectory in HGT1000 from February 8th to February 12th in 1934. To change trajectory from February 9th, we look for analogue of the current day and its following day D+1. This is done to ensure that the two initial states are similar (high pressure system located over France on February 9th 1934 and on its analogue, January 28th 1921) and that the main features move in similar directions (high pressure system heading South-East on both February 10th 1934 and January 29th 1921). The same method is applied to jump from January 30th in 1921 to February 12th in 1925. One can then compare the large differences between the synoptic weather situation of the last day of the observed atmospheric sequence (February 12th 1934) and the one of

the “constructed atmospheric trajectory” (February 12th 1925). Two additional criteria are necessary to guarantee relevant and correct sequences of days. First, the reconstructed series must rest on a given calendar sequence and the analogue to change trajectory must no move too far from it or the method might quickly mix summer and winter days. To ensure this condition, we look for the 5 best HGT1000 analogues on a +/-15 days seasonal window and we allow a jump in trajectories only if the difference between the target calendar day and the one used analogue date is lower than 15 days. The combination of the two constraints gives a maximum lag time of 30 days between the target calendar day and the analogue used in the reconstructed series. The decision tree applied every day and used to decide whether to change trajectory or not is presented in Figure A. 9.

The frequency of trajectory change has also been calibrated in order to guarantee a good climatology of the weather situations. To do so, we analysed the frequency and durations of each of the 9 weather types proposed by Philipp et al. (2010) in the observed synoptic situations and in the reconstructed one for different probability of trajectory change (from 1 to 10 days). The results showed that this probability does not impact the frequency of weather types but that a minimum probability of 7 days is necessary to generate weather types that exhibit the wide and relevant range of persistency (Figure A. 10). Thus, our methods used a probability of 1/7 to decide whether to change trajectory or not.

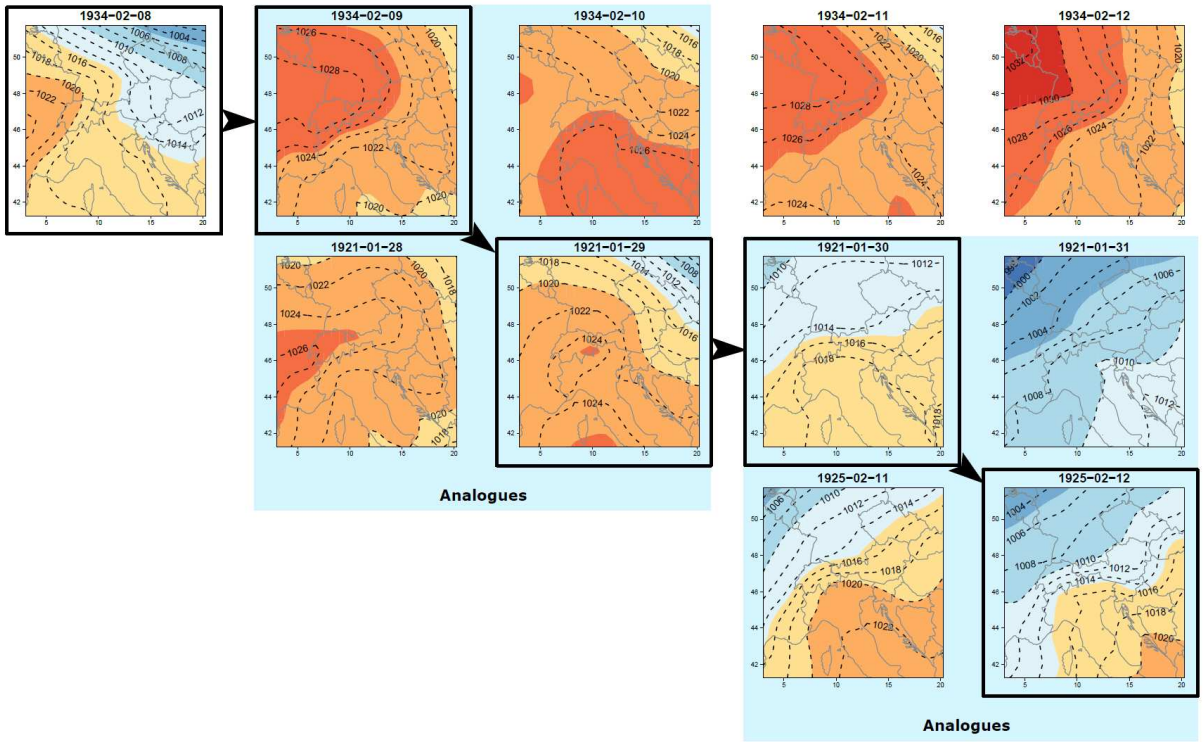


Figure A. 8 Construction of a new atmospheric trajectory from an observed synoptic weather sequence. Each sub-figure presents the geopotential at 1000hPa on the domain of interest. The black squares and arrows give the new atmospheric trajectory and the blue shading highlight the two-day analogue that help “changing of atmospheric direction”.

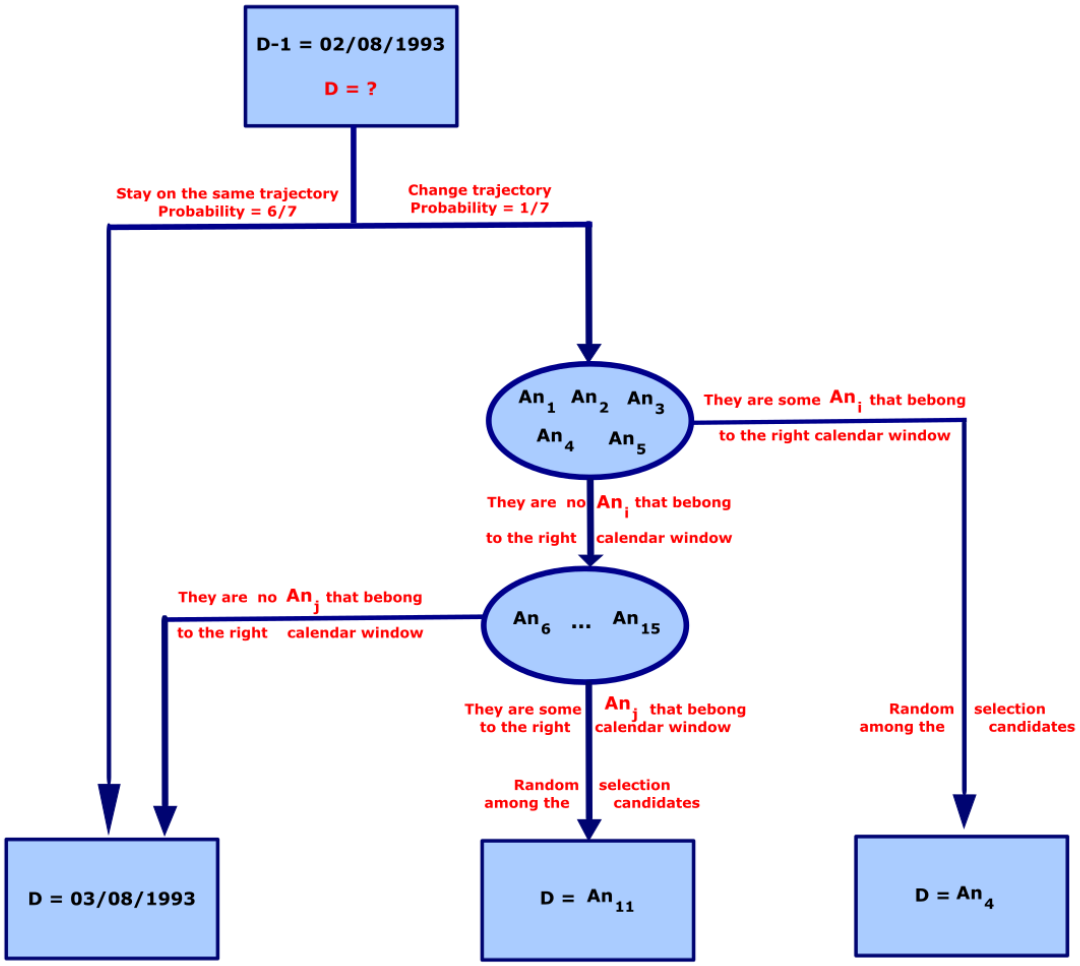


Figure A. 9 Decision tree and the different tests taking part in the construction of a new atmospheric trajectory. Day D-1 is already defined and we here explain the different option for the choice of day D. "An" stands for the analogue days.

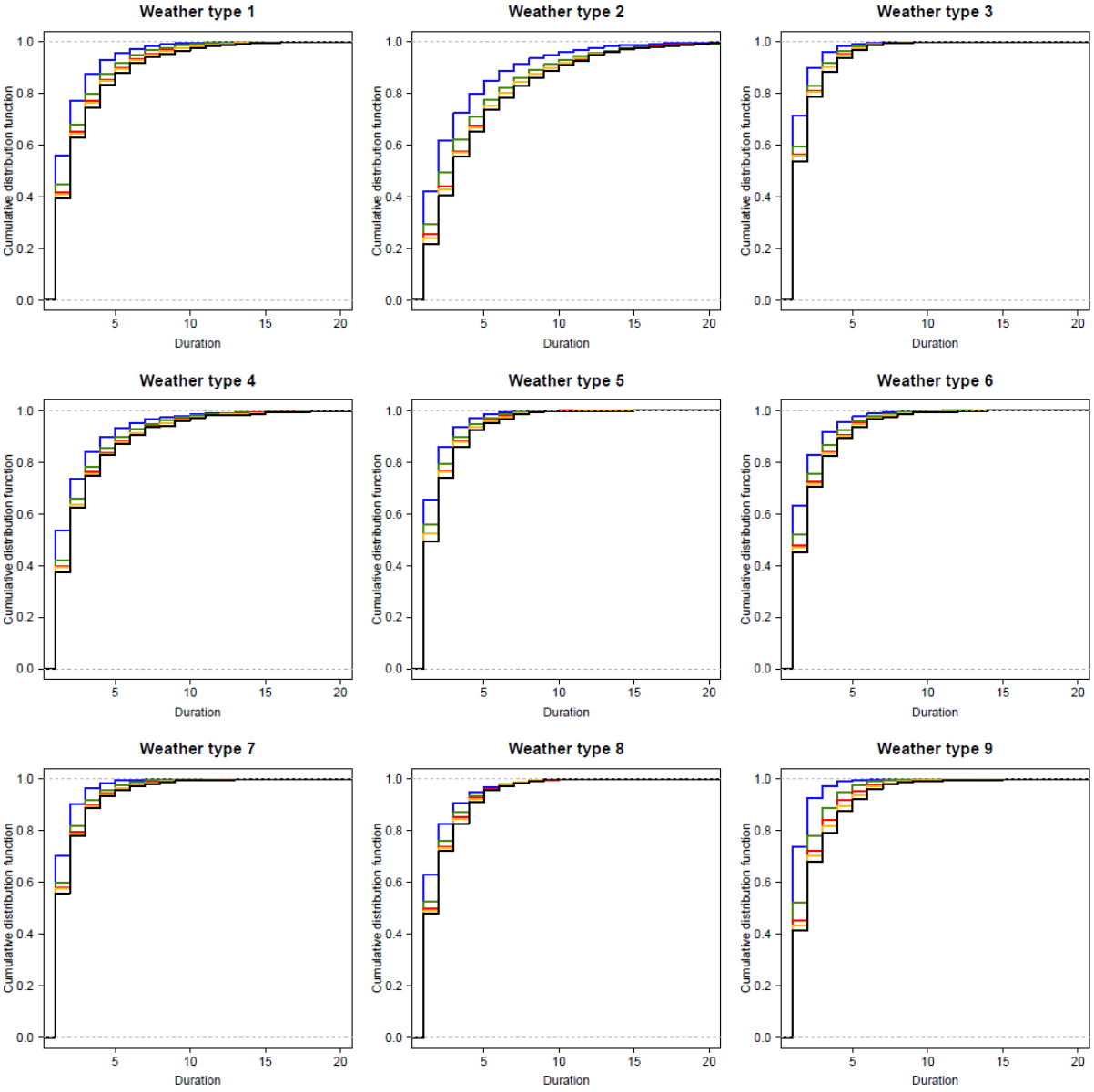


Figure A. 10 Persistency (in days) of each of the 9 weather types defined for Switzerland (Philipp et al., 2010) in observations (black) and in 1000-yr long time series of atmospheric trajectories for jump probabilities of 1day (blue), 3 days (green), 7 days (red) and 10 days (orange).

10.4.2 Spatial downscaling from regional precipitation to local scenarios on the sub-catchments

We present here the functioning of the spatial disaggregation model that converts the regional precipitation and temperature scenarios into local scenarios over the 89 sub-catchments. To do so, we use daily observed spatial structures available from 1930 to 2014.

Two main conditions need to be fulfilled to get a satisfactory spatial downscaling:

- The local scenarios for precipitation must be physically realistic, meaning that extremely large values must not be generated unless the daily regional precipitation also corresponds to an extreme event.

- As the final time series will be at hourly time step, we want to avoid temperature jumps from D-1 at 11pm to D 0am that would be too large and unrealistic too.

For each target day and regional scenario, we always give priority to the spatial structures associated to the 30 analogue dates. However, it happens that no analogue fulfils the two conditions mentioned previously. In this case, we use a set of reference structures that take over the spatial downscaling. This set of structures is made of observed structures that present a large diversity precipitation and temperature characteristics. For each month, and each precipitation class such as: $P_{class}/\ 1.5 \leq P_{regional} \leq P_{class} \times 1.5 \mid \quad P_{class} \in [0,\ 0.1,\ 0.2,\ 0.3,\ 0.4,\ 0.5,\ 0.8,\ 1,\ 1.5,\ 2,\ 3,\ 5,\ 10,\ 15, 20]$, we randomly selected 100 days and their associated precipitation and temperature spatial structures. In addition, a last regional precipitation class containing to the 50 maximum observed precipitation events for each month has been added to help downscaling extreme regional precipitation.

The first condition, preventing from generating unrealistic local precipitation is guaranteed by defining and using a threshold depending on the regional precipitation value of each scenario:

$$\mu(P_{sce}) = 1.5\ \times 10^{-4} \times P_{Sce}^2.$$

An observed spatial structure for precipitation is considered suitable if its regional precipitation value is such as $PI = \frac{P_{Sce}}{P_{obs}} \leq \mu(P_{sce})$.

Similarly, for the condition focusing on temperature jump, a threshold of 6°C from D-1 at 11pm to D at 0am is used. This value corresponds to the maximum change in regional temperature at this time of the day existing in the observed time series.

The decision tree and the different tests taking part in the choice of a suitable spatial structure for both local predictors are presented in Figure A. 11. The 30 analogue scenarios are first tested on the regional precipitation condition and then on the temperature jump one. If there is more than one analogue fulfilling the two conditions, the analogue leading to the minimum temperature jump to the previous day is selected.

When there is no suitable analogue, the reference structures set is used. We allow looking for spatial structure that belongs to a 3-month seasonal window centered on the month of the current target day. The condition on the temperature jumps is only applied if there are some reference structures fulfilling the condition of the PI coefficient, i.e. if we do not deal with an extreme precipitation event.

Indeed, as these events are the main target of the EXAR project, we do not want to introduce too many constraints and we only focus on picking the most suitable spatial structure.

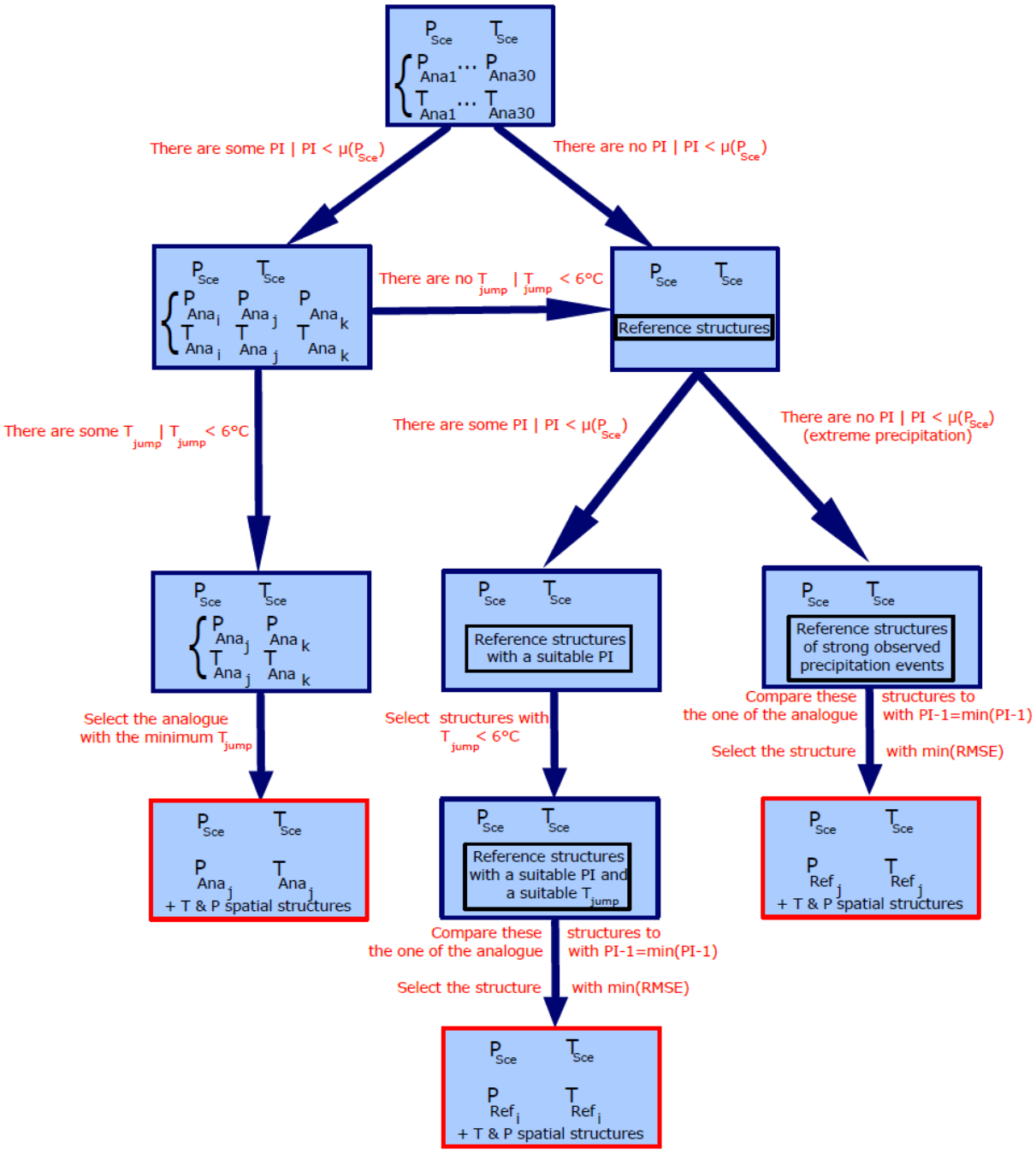


***Figure A. 11* Decision tree and the different tests for the choice of a suitable spatial structure for the downscaling on the regional precipitation and temperature scenarios. $PI = \frac{P_{Sce}}{P_{obs}}$ where P_{sce} is the regional precipitation scenarios to be downscaled, and P_{obs} =P_{Ana} regional precipitation of an analogue or $= P_{Ref}$ regional precipitation associated to a reference structure. $\mu(P_{sce}) = 1.5 \times 10^{-4} \times P_{Sce}^2$ moving threshold for the PI coefficient. T_{jump} temperature jump from D-1 at 11pm to D 0am.**

An additional and final test is applied when using the reference structure. Indeed, the conditions on PI and T_{jump} do not guarantee that the spatial structure of precipitation agrees well with the synoptic meteorological situation. To prevent our downscaling method to lead to irrelevant structure, we compute the RSME score between the remaining candidates from the reference precipitation

structures and the one of the analogues having the closest PI coefficient to 1. By doing so, we avoid picking a precipitation structures that are completely inverted compared to the ones of the analogues (e.g. precipitation on the western part of the test region in the reference structure versus precipitation on the eastern part on for the analogues).

References

Philipp, A., Bartholy, J., Beck, C., Erpicum, M., Esteban, P., Fettweis, X., Huth, R., James, P., Jourdain, S., Kreienkamp, F. and Krennert, T., 2010. Cost733cat–A database of weather and circulation type classifications. *Physics and Chemistry of the Earth, Parts A/B/C*, *35*(9-12), pp.360-373.

Yiou, P., 2014. Anawege: a weather generator based on analogues of atmospheric circulation. *Geoscientific Model Development*, *7*(2), pp.531-543.

10.5 Weather scenarios – Evaluations

10.5.1 SCAMP versus GWEX

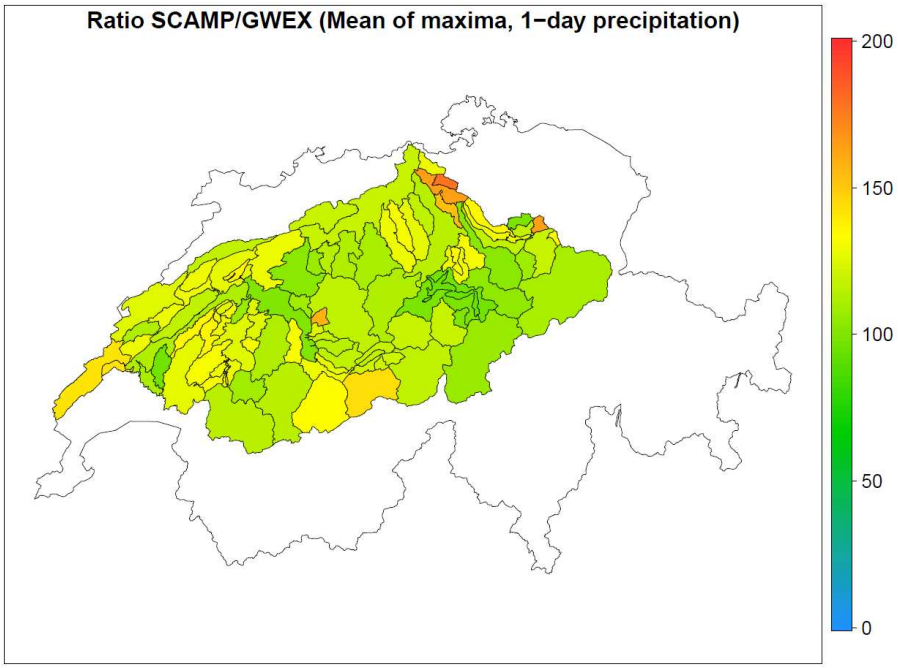


Figure A. 12 SCAMP-GWEX ratio (in percentage) for MAP1d

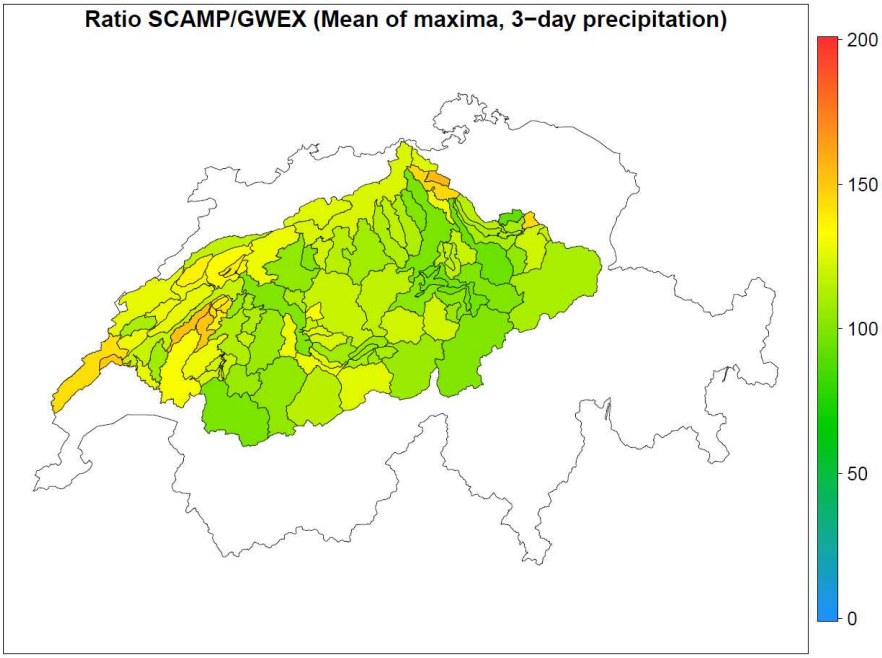


Figure A. 13 Figure B: SCAMP-GWEX ratio (in percentage) for MAP3d

10.5.2 Largest precipitation events with GWEX: Severity maps and space/time dynamics

The precipitation space-time patterns during the largest precipitation events are shown in the following figures: for each event first the severity map is show, then the space/time dynamics of precipitation. The severity map represents the severity of this event at several spatial scales and for several accumulation periods; the space/time dynamics of precipitation during 10 days are composed of 10 maps ("Day 0" is the day of the precipitation peak).

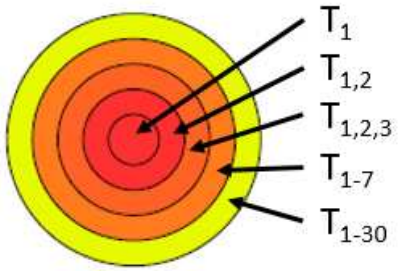
The title of the severity map shows the rank (in descending order) of the precipitation events at the whole catchment scale for several accumulation durations (all ending the day of the precipitation peak):

- P_1[]: rank of the 1-day precipitation during the day of the precipitation peak (“Day 0”)
- P_2[]: rank of the 1-day precipitation during the day before the precipitation peak
- P_3[]: rank of the 1-day precipitation 2 days before the precipitation peak
- $P_{1,2}$[]: rank of the 2-day precipitation
- $P_{1,2,3}$[]: rank of the 3-day precipitation
- $P_{1\text{-}7}$[]: rank of the 7-day precipitation
- $P_{1\text{-}30}$[]: rank of the 30-day precipitation
- In addition, the fictive date of the precipitation peak is indicated on the third line

The severity of the simulated event is then represented with a set of severity-board. Each severity-board gives the return period of precipitation simulated for this event when cumulated over several durations (all ending the day of the precipitation peak):

- 1 day (center of the severity-board)
- 2 days
- 3 days
- 7 days
- 30 days (edge of the severity-board)

For instance, if T_1, $T_{1,2}$, $T_{1,2,3}$, $T_{1\text{-}7}$, and $T_{1\text{-}30}$ are the return periods associated with the precipitation P_1, $P_{1,2}$, $P_{1,2,3}$, $P_{1\text{-}7}$, and $P_{1\text{-}30}$, respectively



The color scale for the severity-boards is given in the right hand side of the figure. Severity-boards are given for precipitation simulated over different spatial scales:

- local precipitation at each of the 105 simulation stations (small circles)
- mean areal precipitation over each of the 5 main sub-catchments (medium circles)
- whole catchment (large circle)

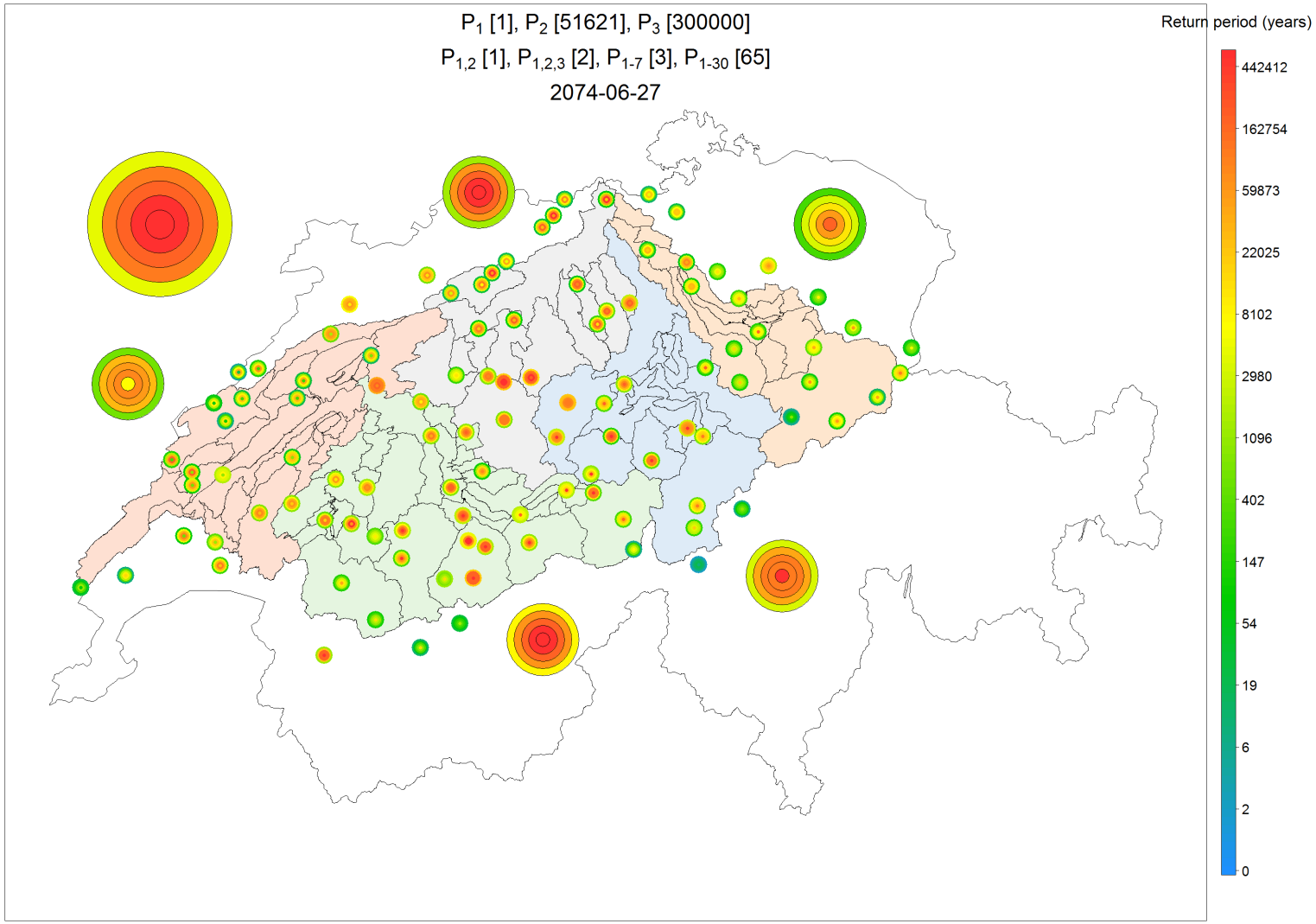


Figure A. 14 Severity map representing the severity at several spatial scales and for several accumulation periods for the first 1-day precipitation event.

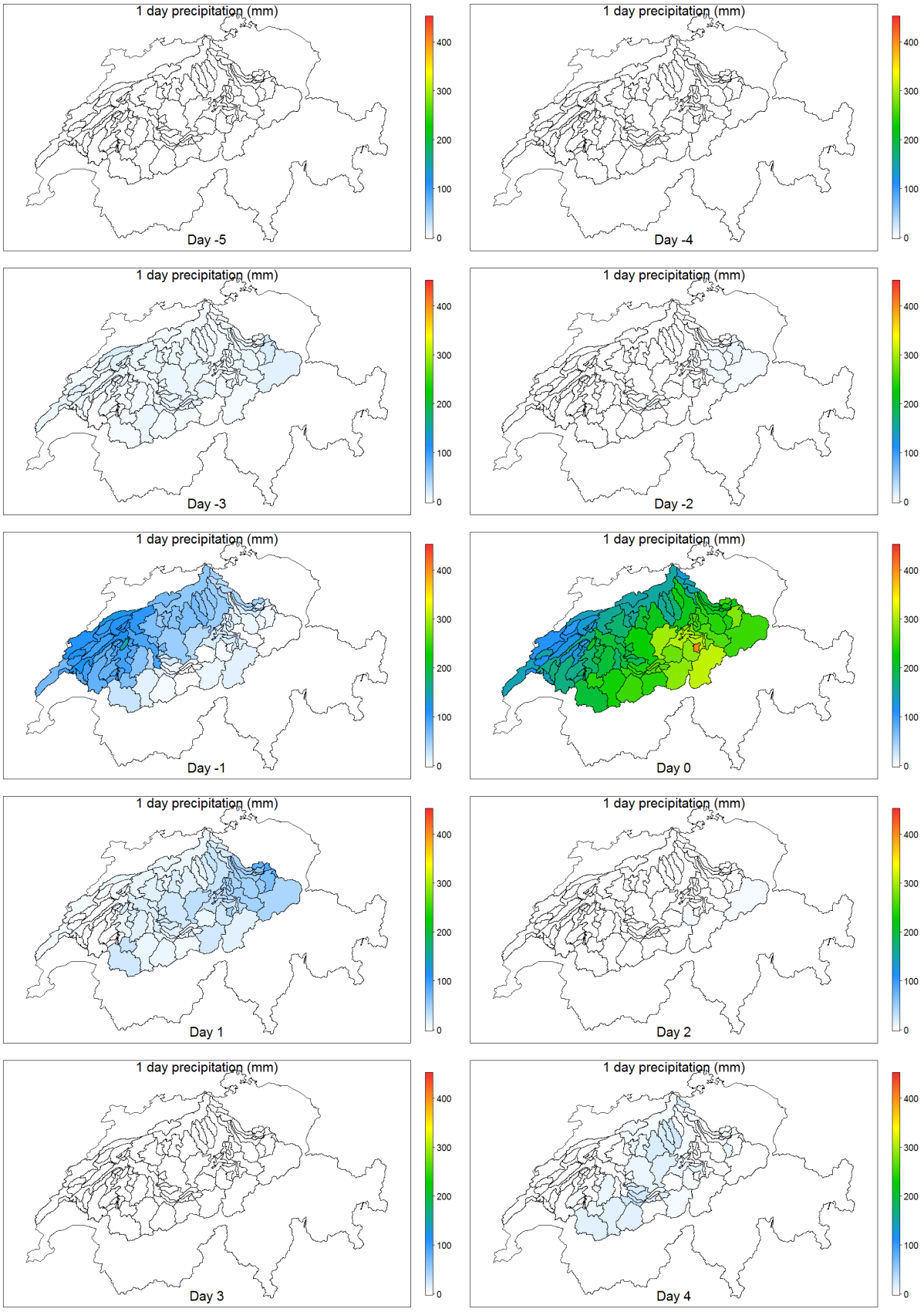


Figure A. 15 Space/time dynamics of precipitation during 10 days for the first precipitation event.

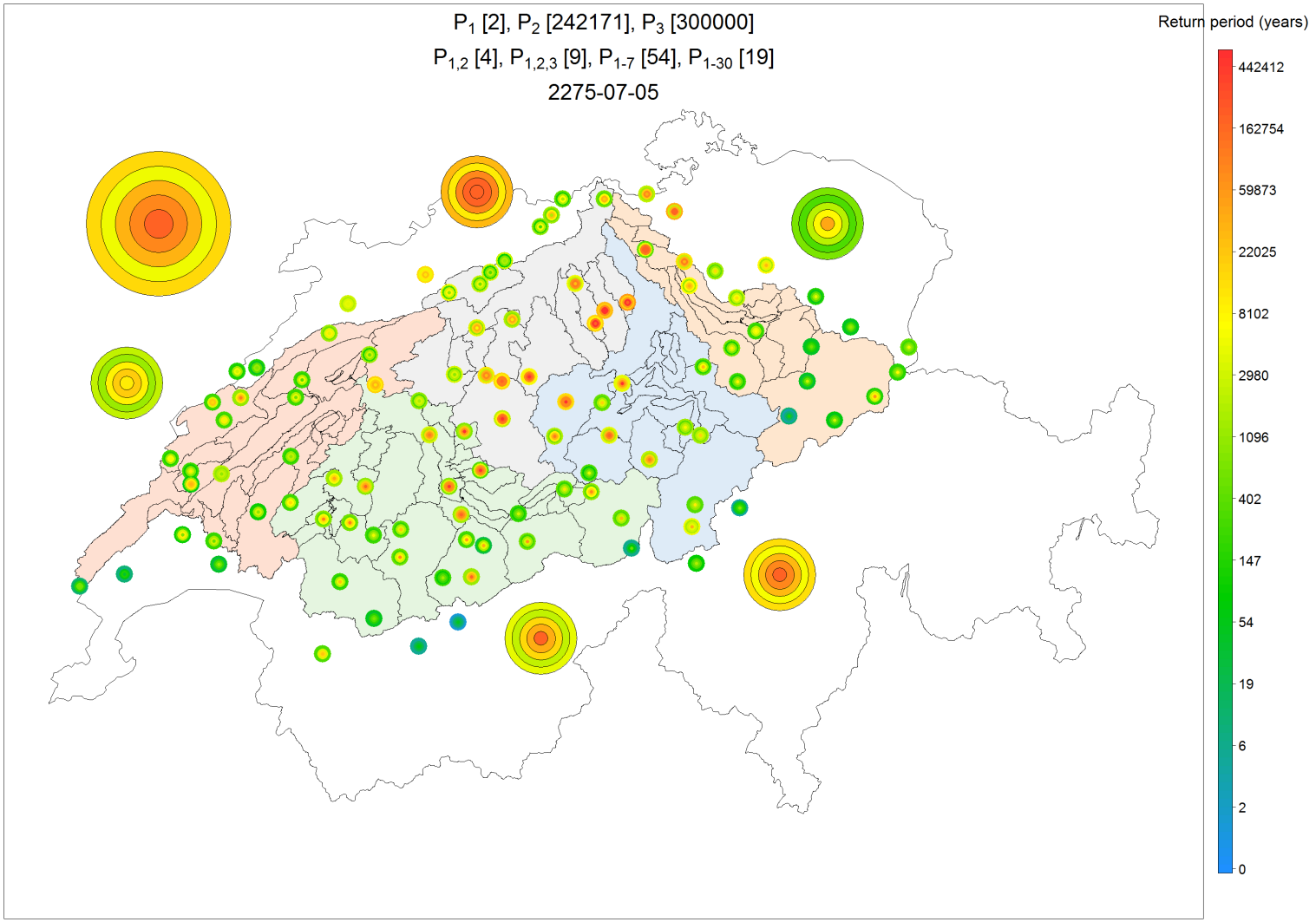


Figure A. 16 Severity map representing the severity at several spatial scales and for several accumulation periods for the 2nd 1-day precipitation event

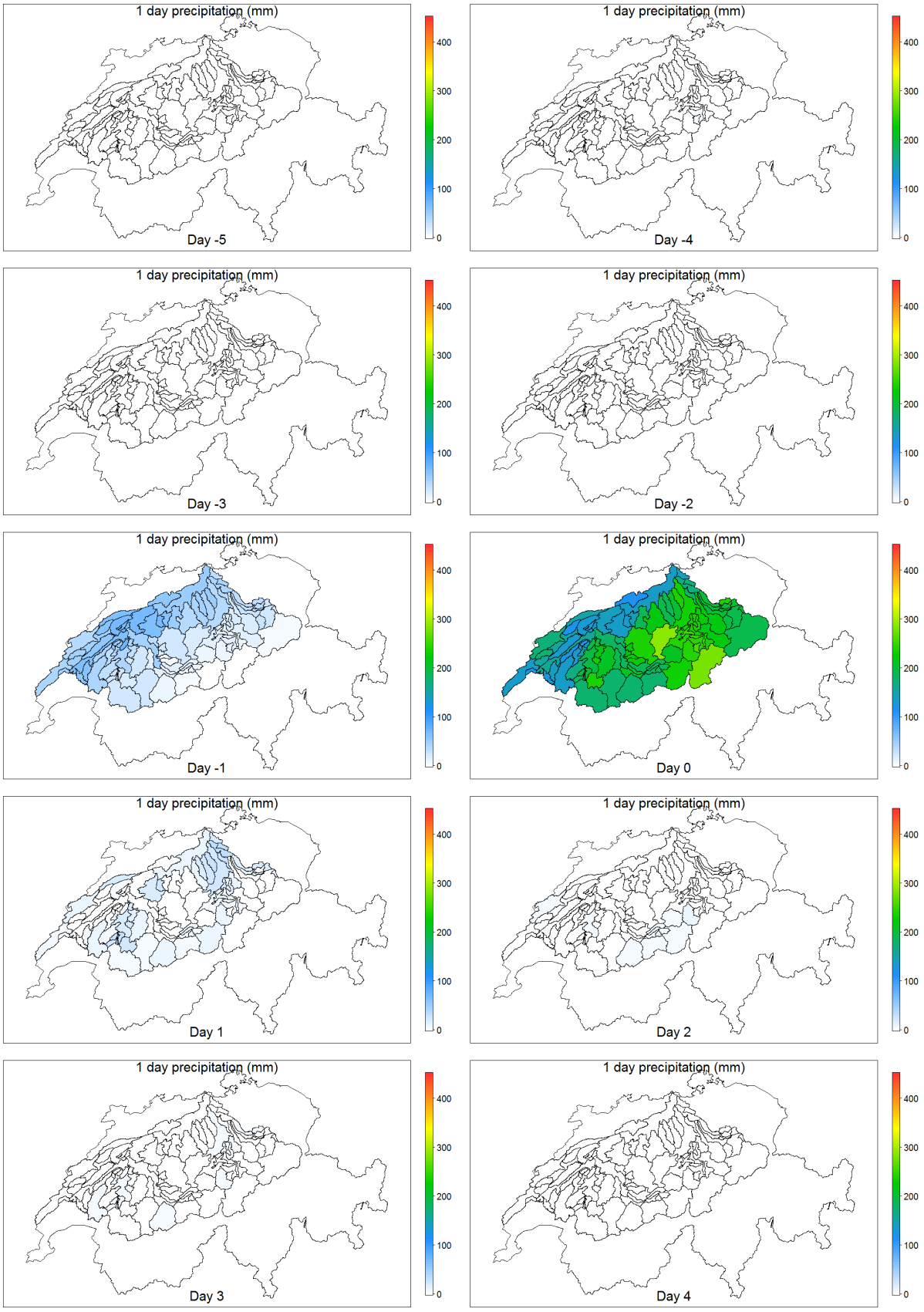


Figure A. 17 Space/time dynamics of precipitation during 10 days for the 2nd 1-day-precipitation event.

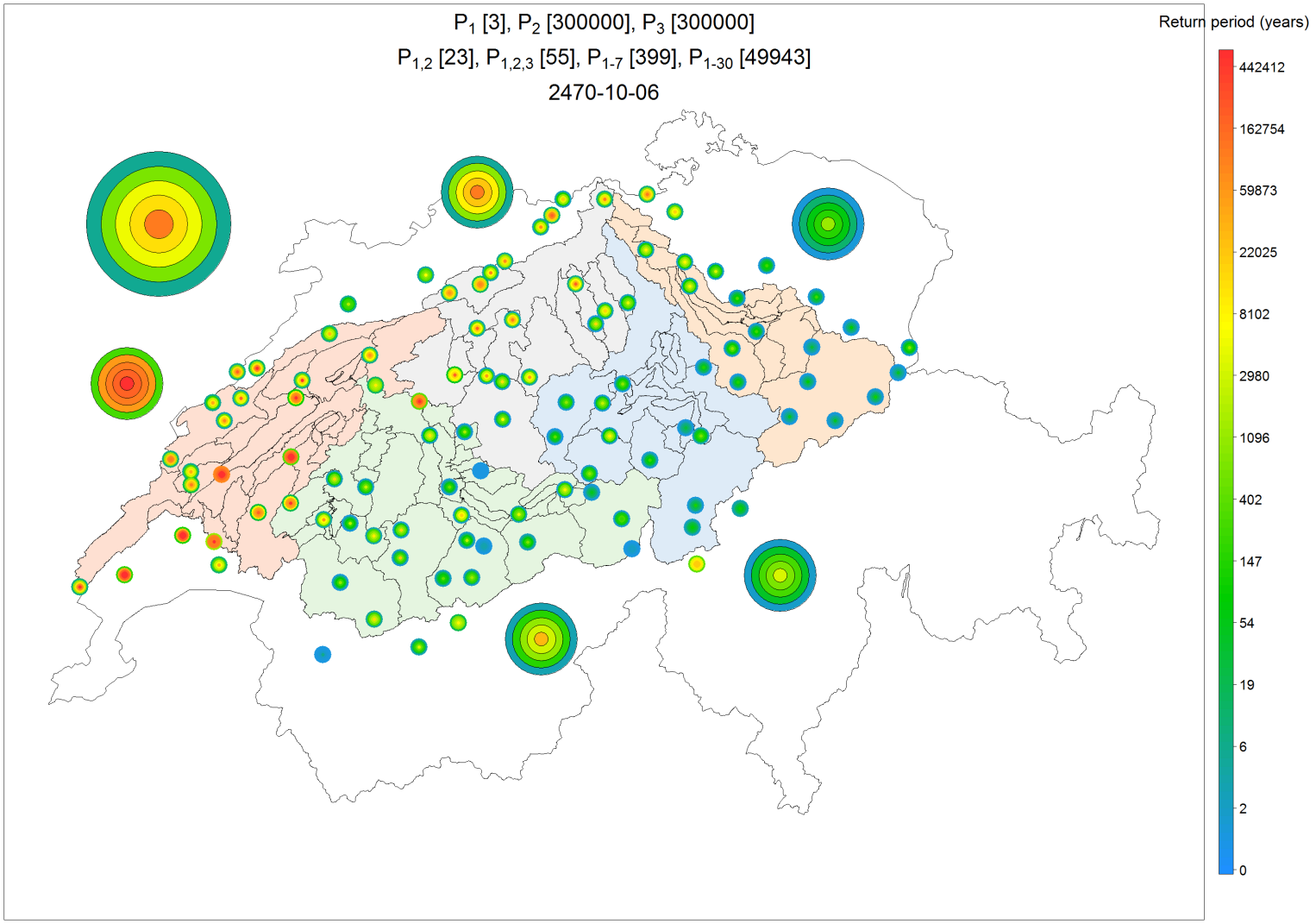


Figure A. 18 Severity map representing the severity at several spatial scales and for several accumulation periods for the 3rd 1-day precipitation event

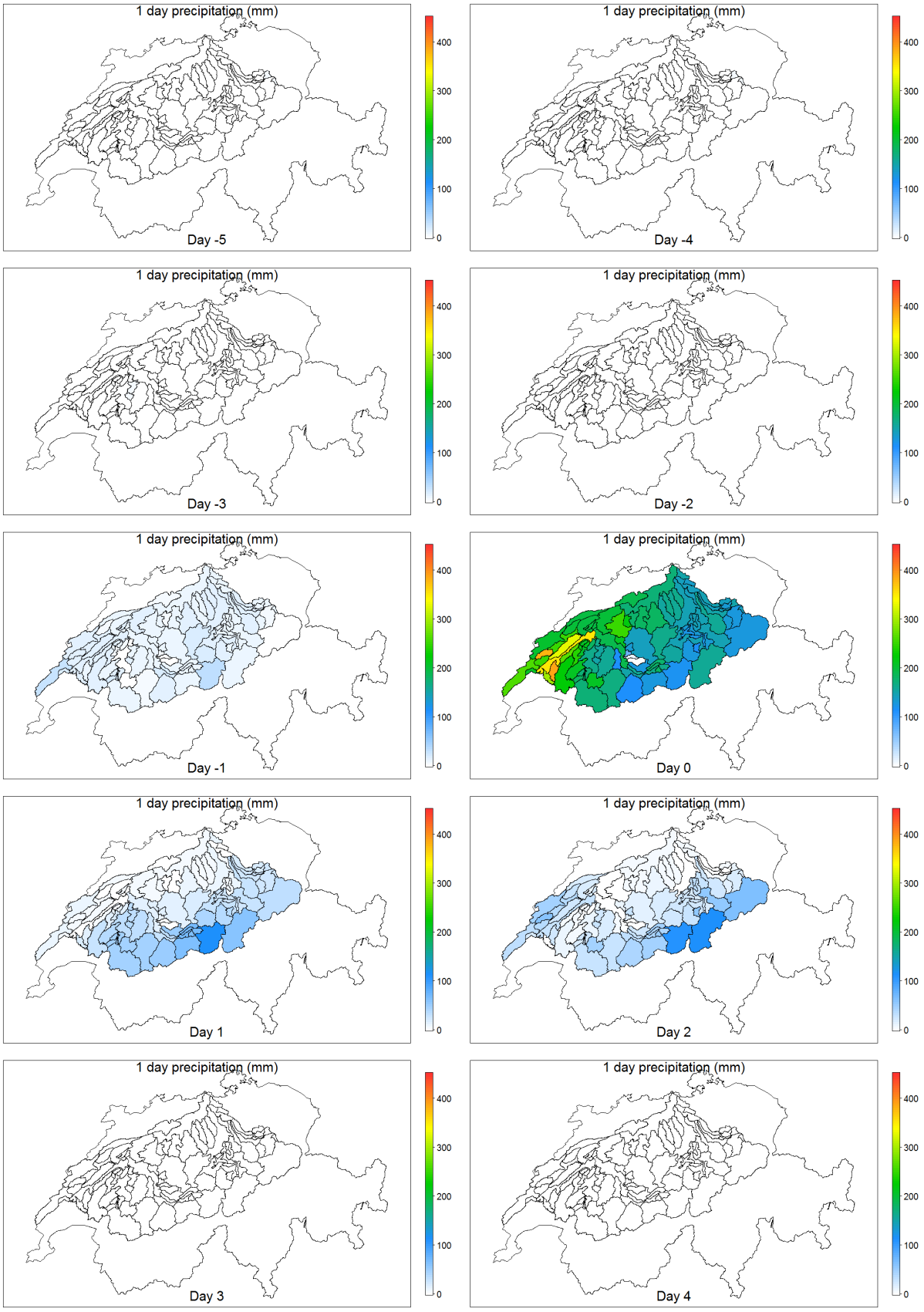


Figure A. 19 Space/time dynamics of precipitation during 10 days for the 3rd 1-day-precipitation event

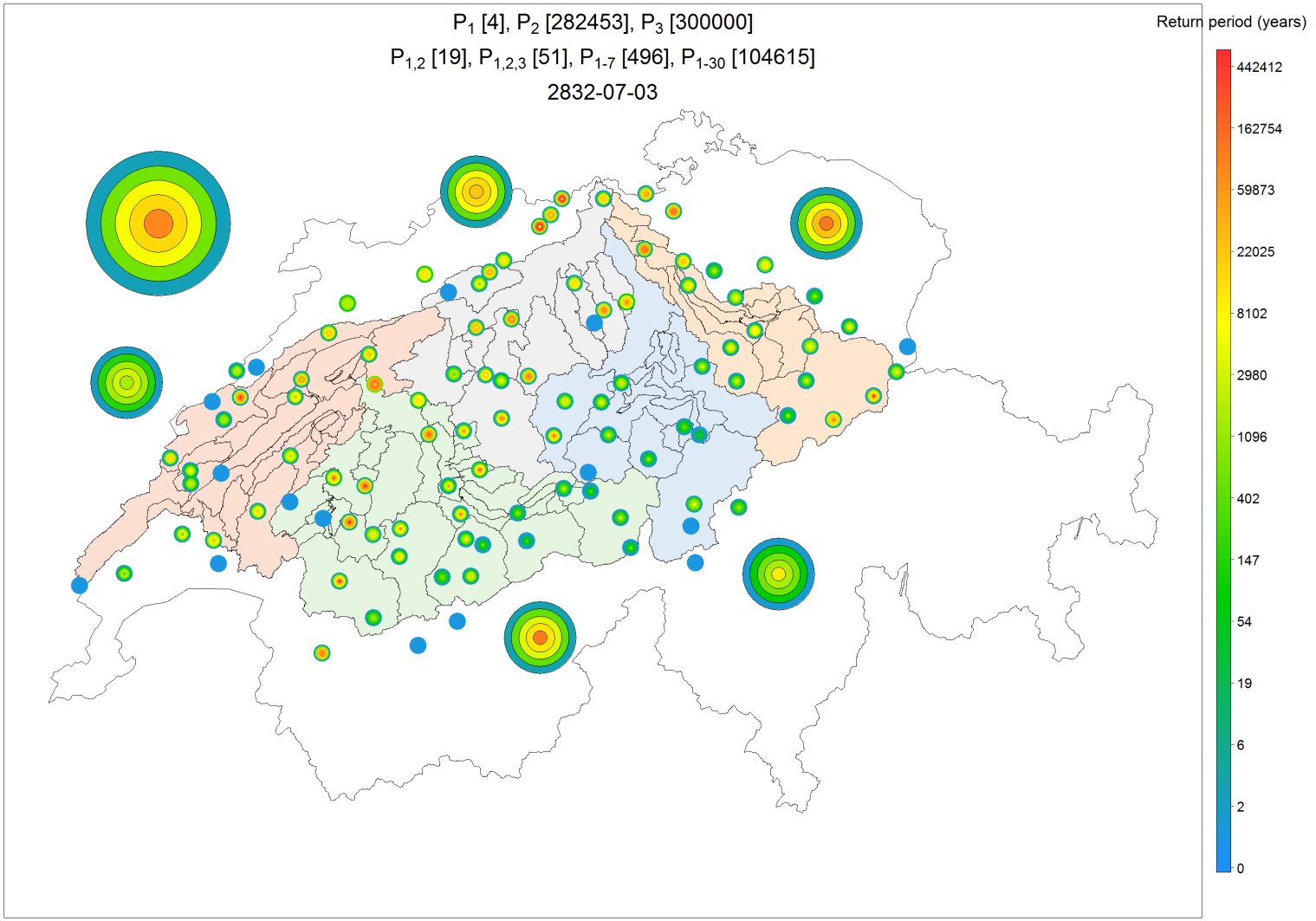


Figure A. 20 Severity map representing the severity at several spatial scales and for several accumulation periods for the 4th 1-day precipitation event

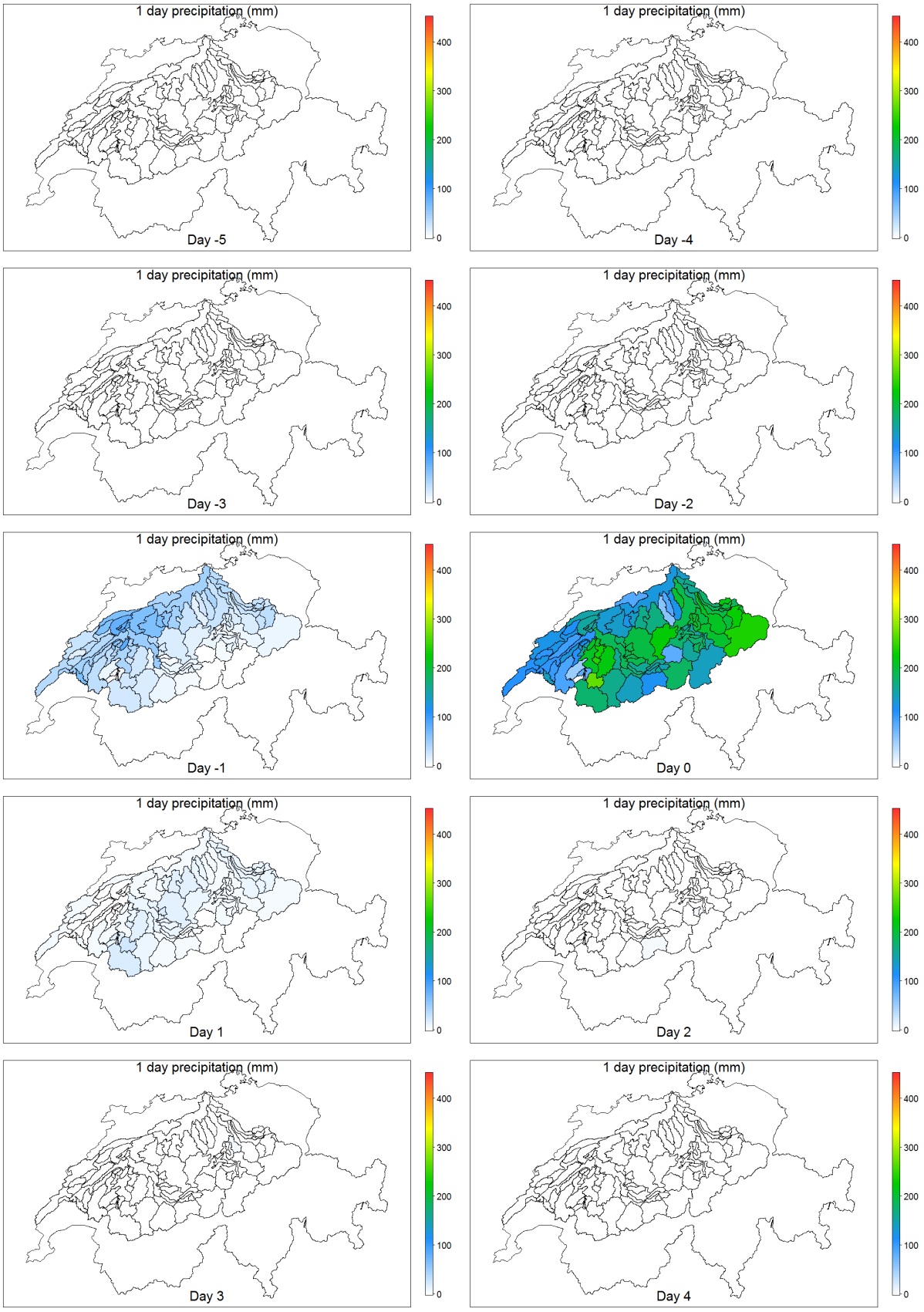


Figure A. 21 Space/time dynamics of precipitation during 10 days for the 4th 1-day-precipitation event

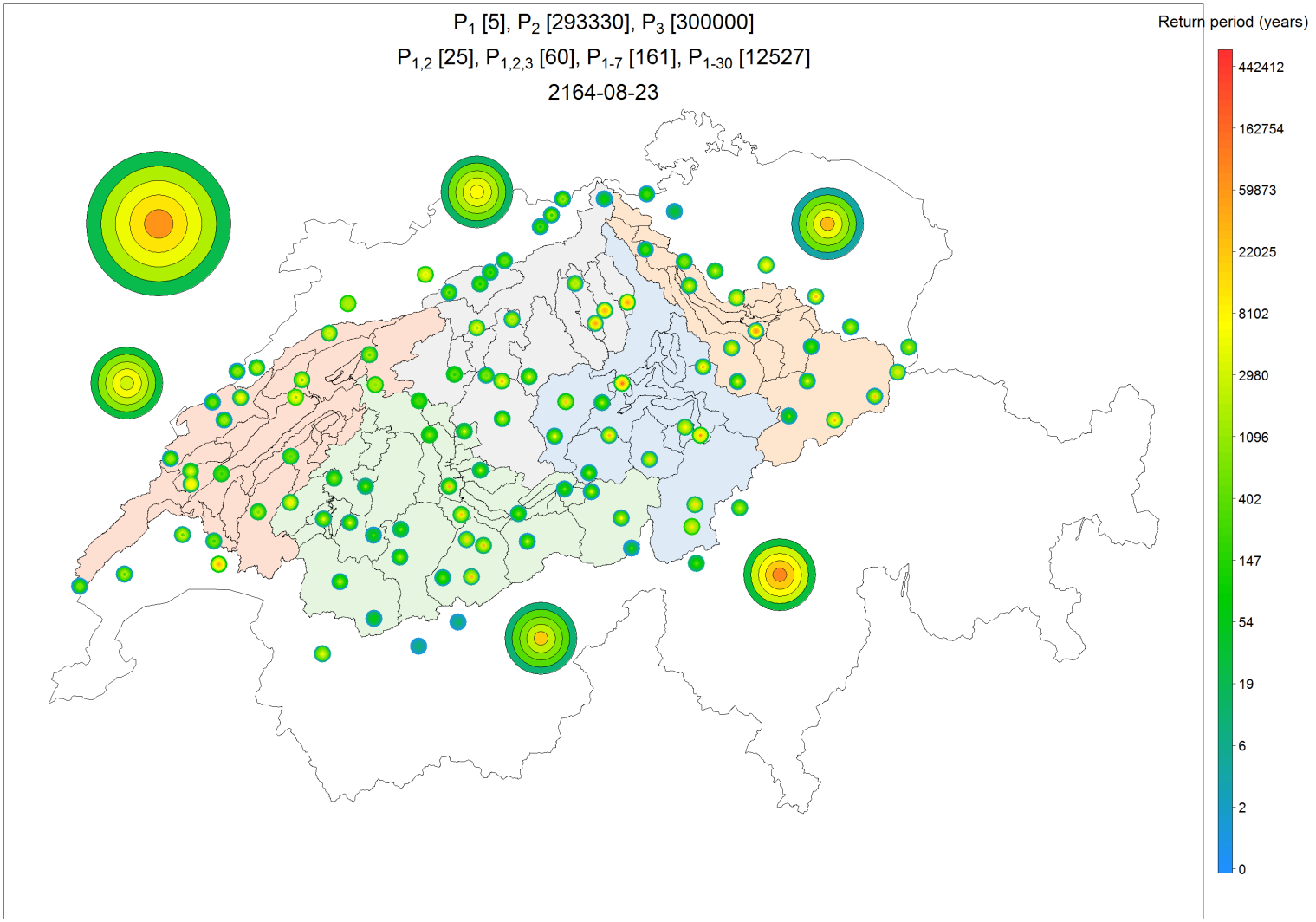


Figure A. 22 Severity map representing the severity at several spatial scales and for several accumulation periods for the 5th 1-day precipitation event

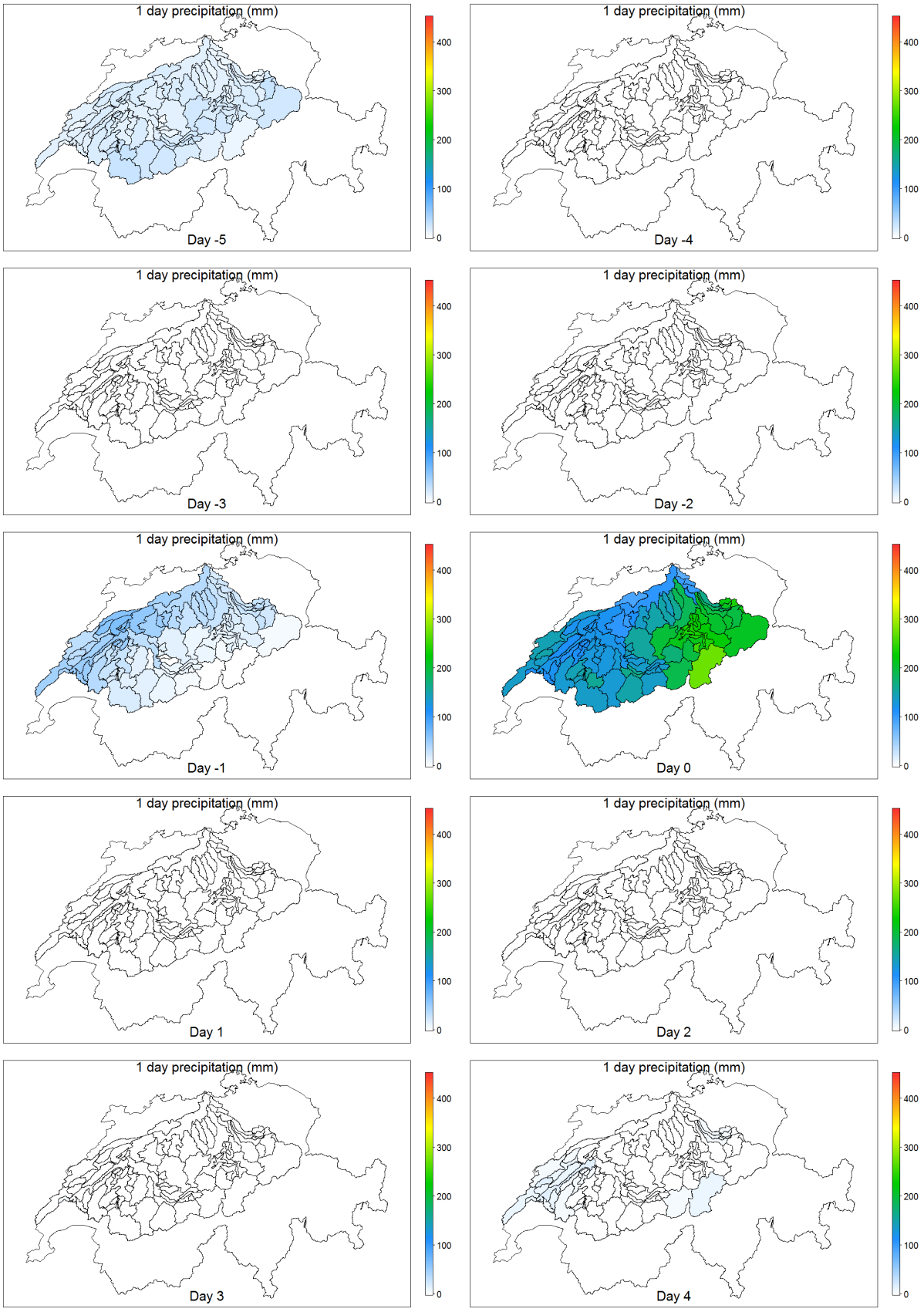


Figure A. 23 Space/time dynamics of precipitation during 10 days for the 5th 1-day-precipitation event.

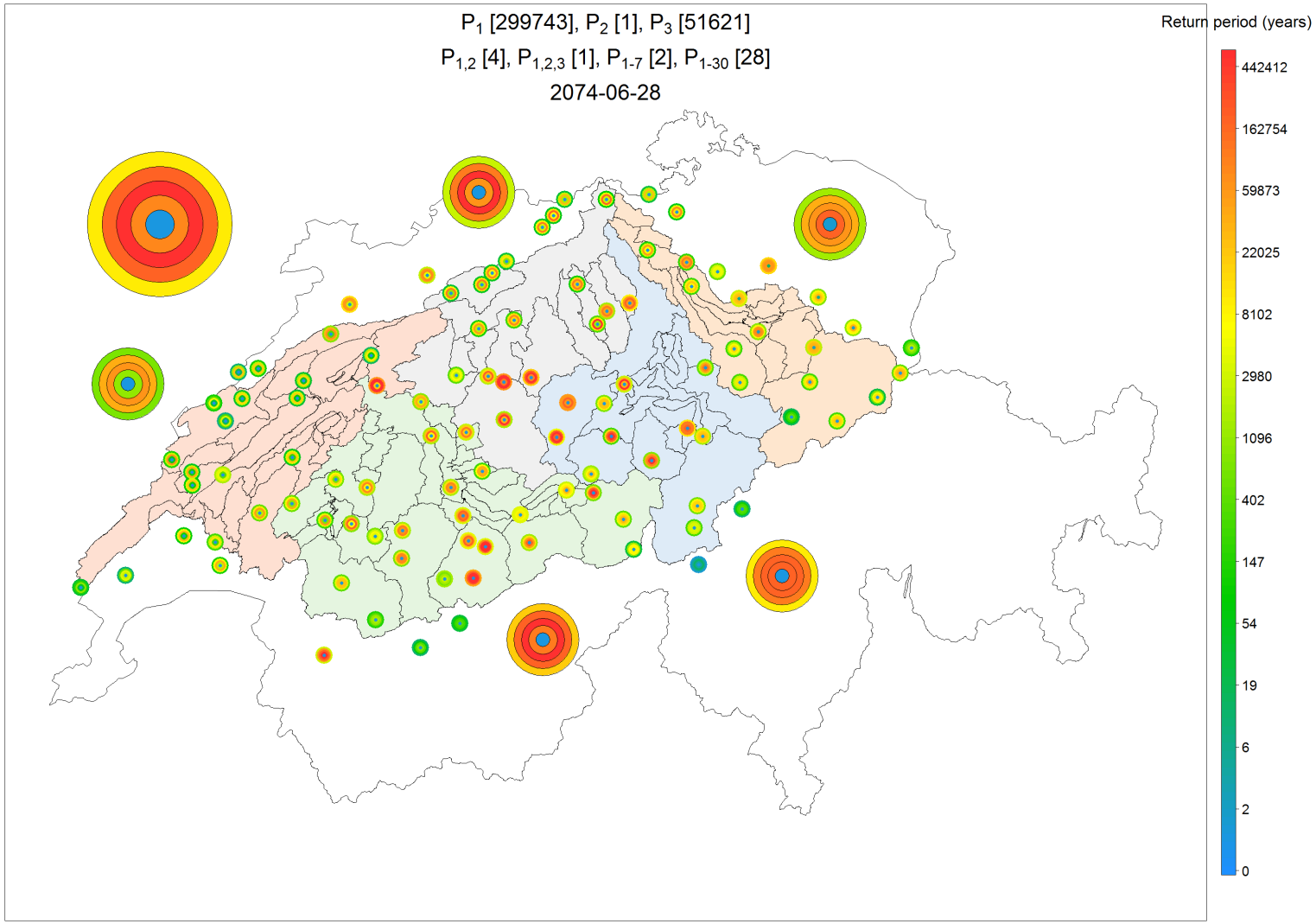


Figure A. 24 Severity map representing the severity at several spatial scales and for several accumulation periods for the 1st 3-day precipitation event.

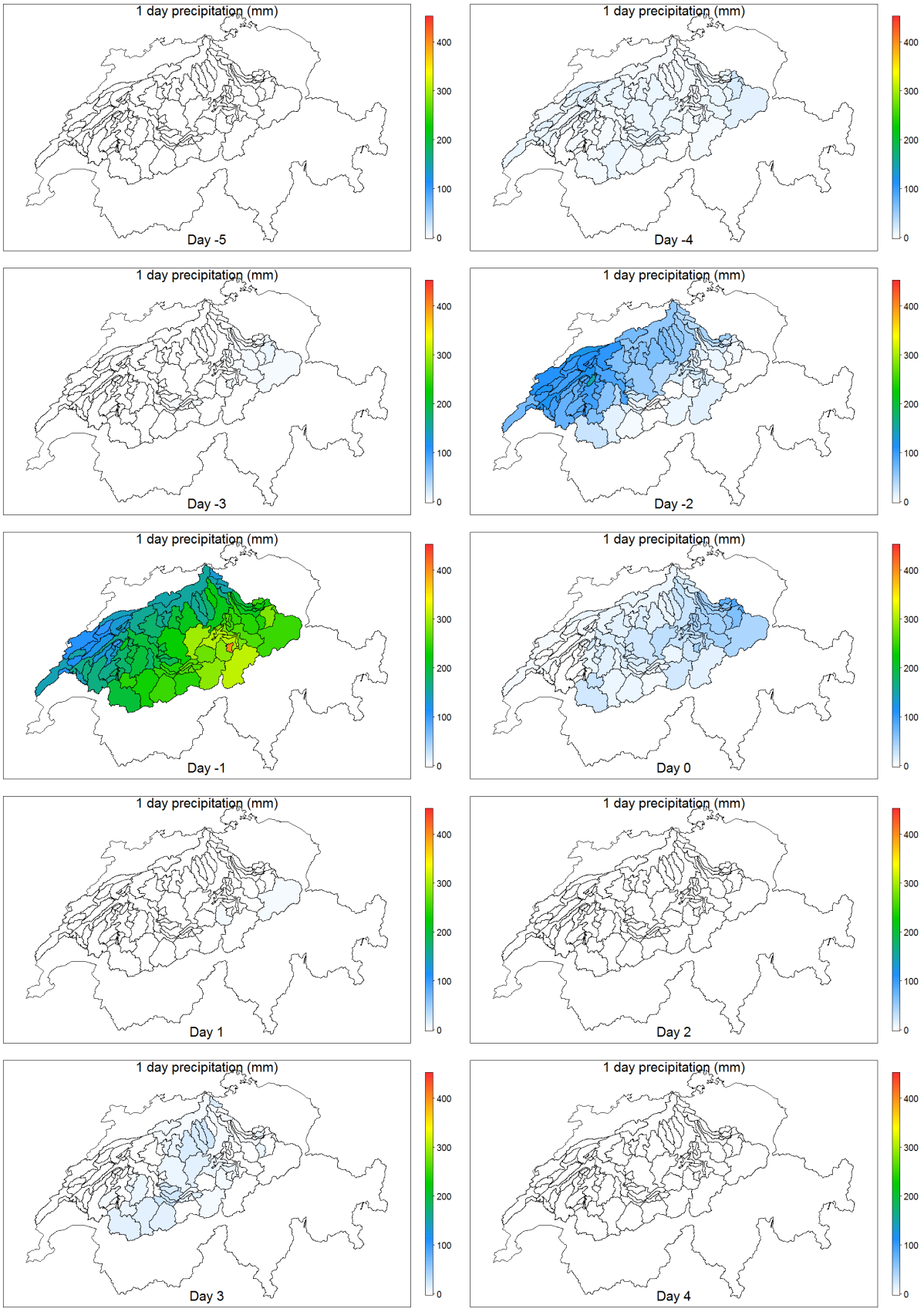


Figure A. 25 Space/time dynamics of precipitation during 10 days for the 1st 3-day-precipitation event.

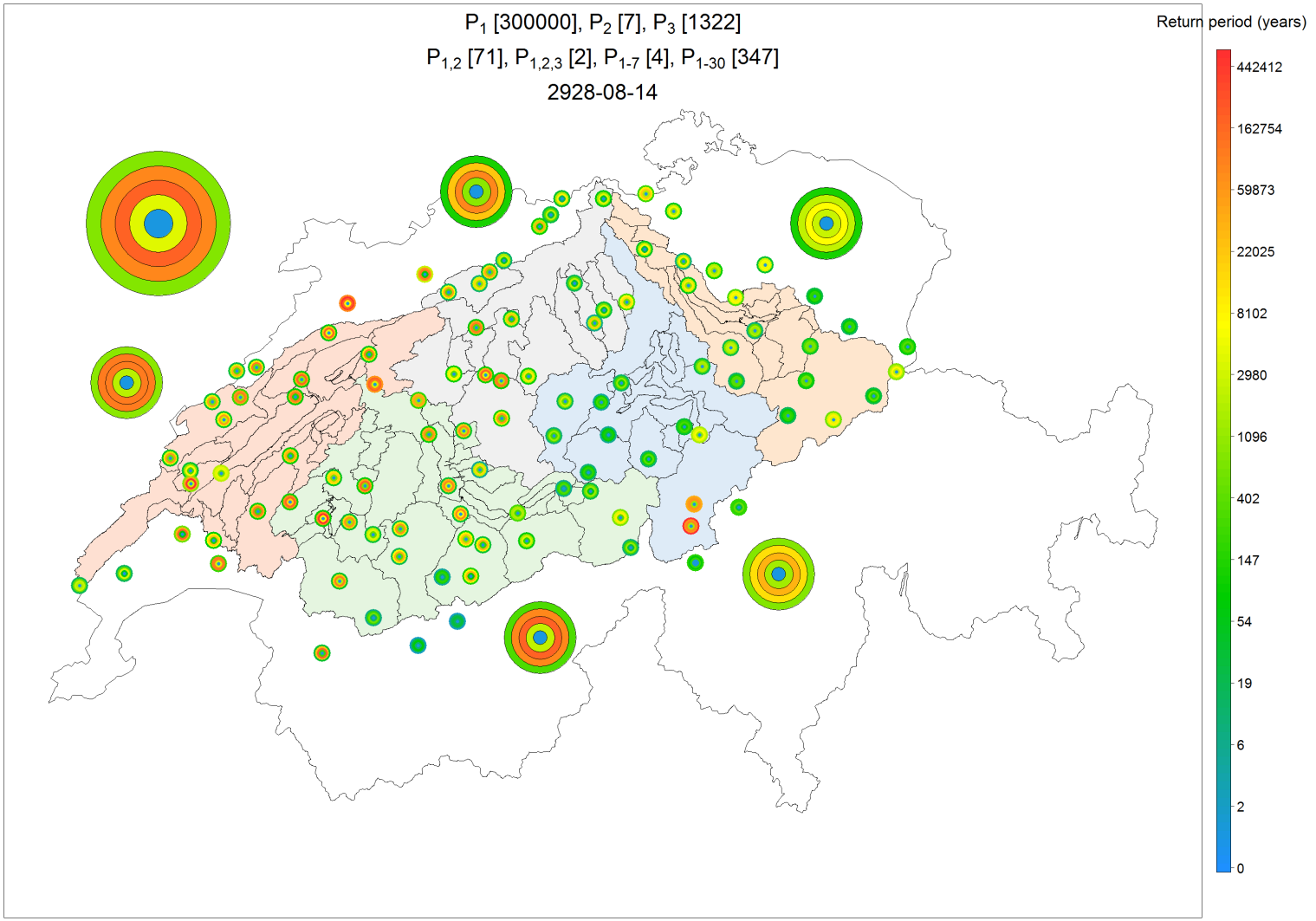


Figure A. 26 Severity map representing the severity at several spatial scales and for several accumulation periods for the 2nd 3-day precipitation event.

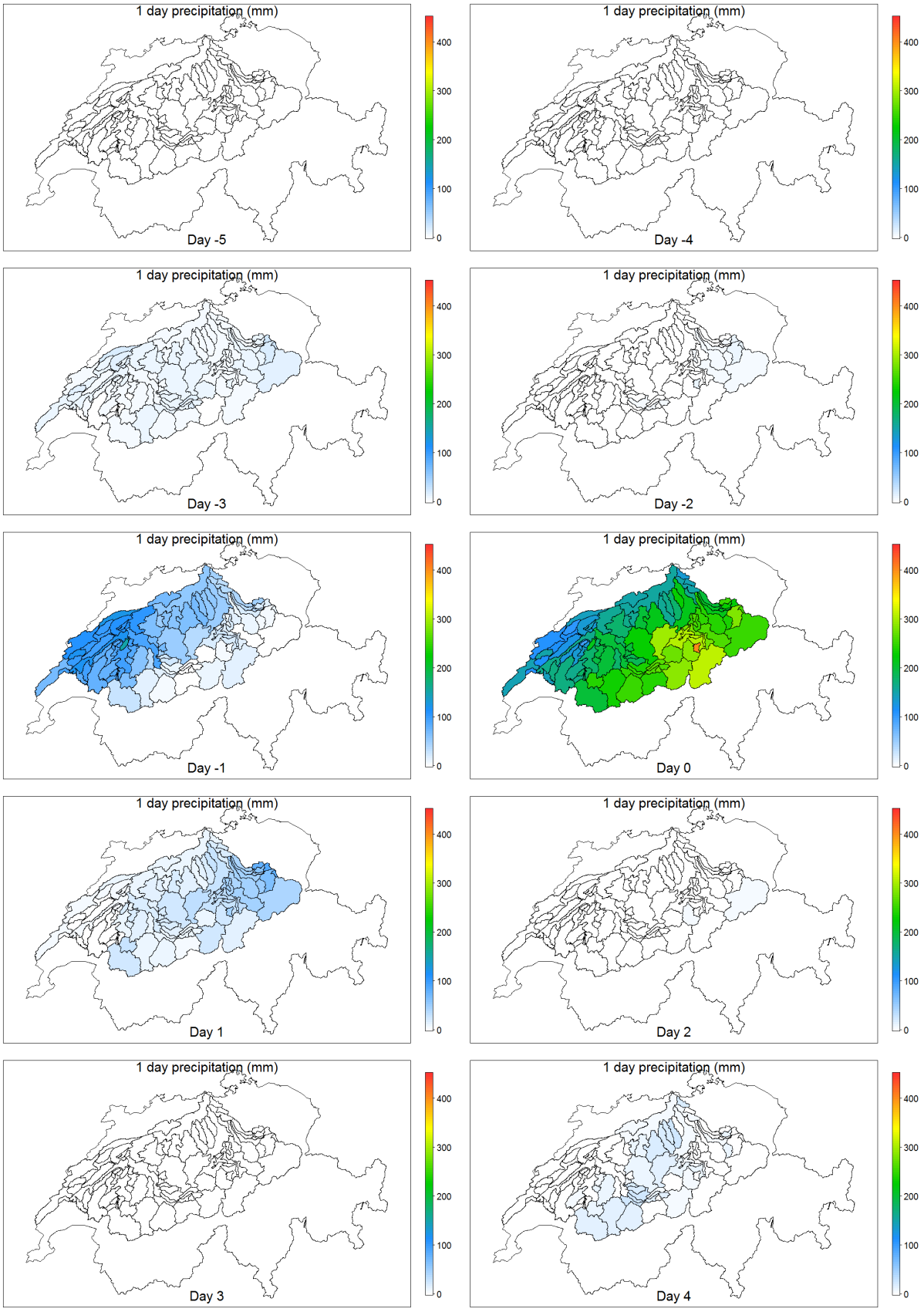


Figure A. 27 Space/time dynamics of precipitation during 10 days for the 2nd 3-day-precipitation event.

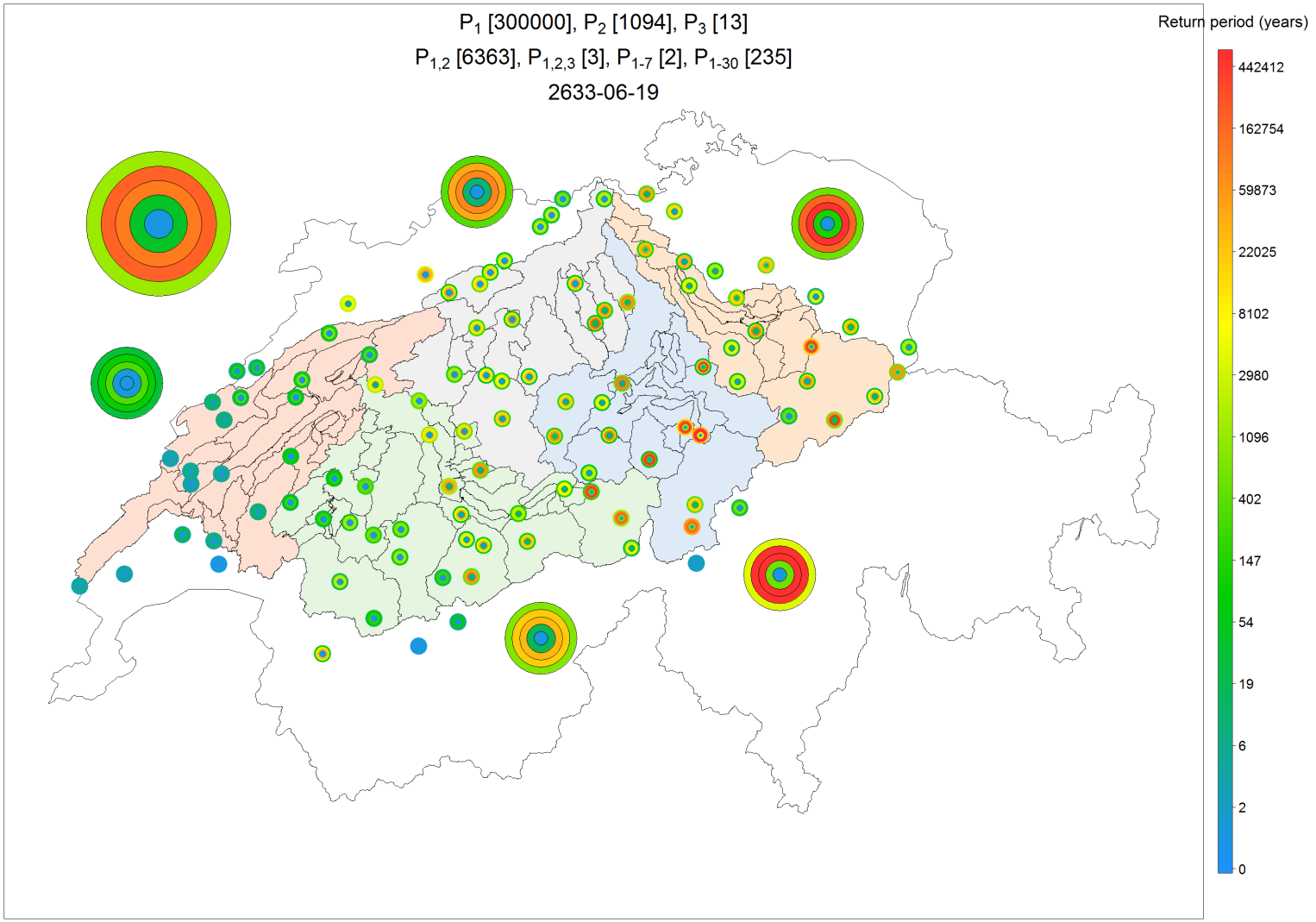


Figure A. 28 Severity map representing the severity at several spatial scales and for several accumulation periods for the 3rd 3-day precipitation event.

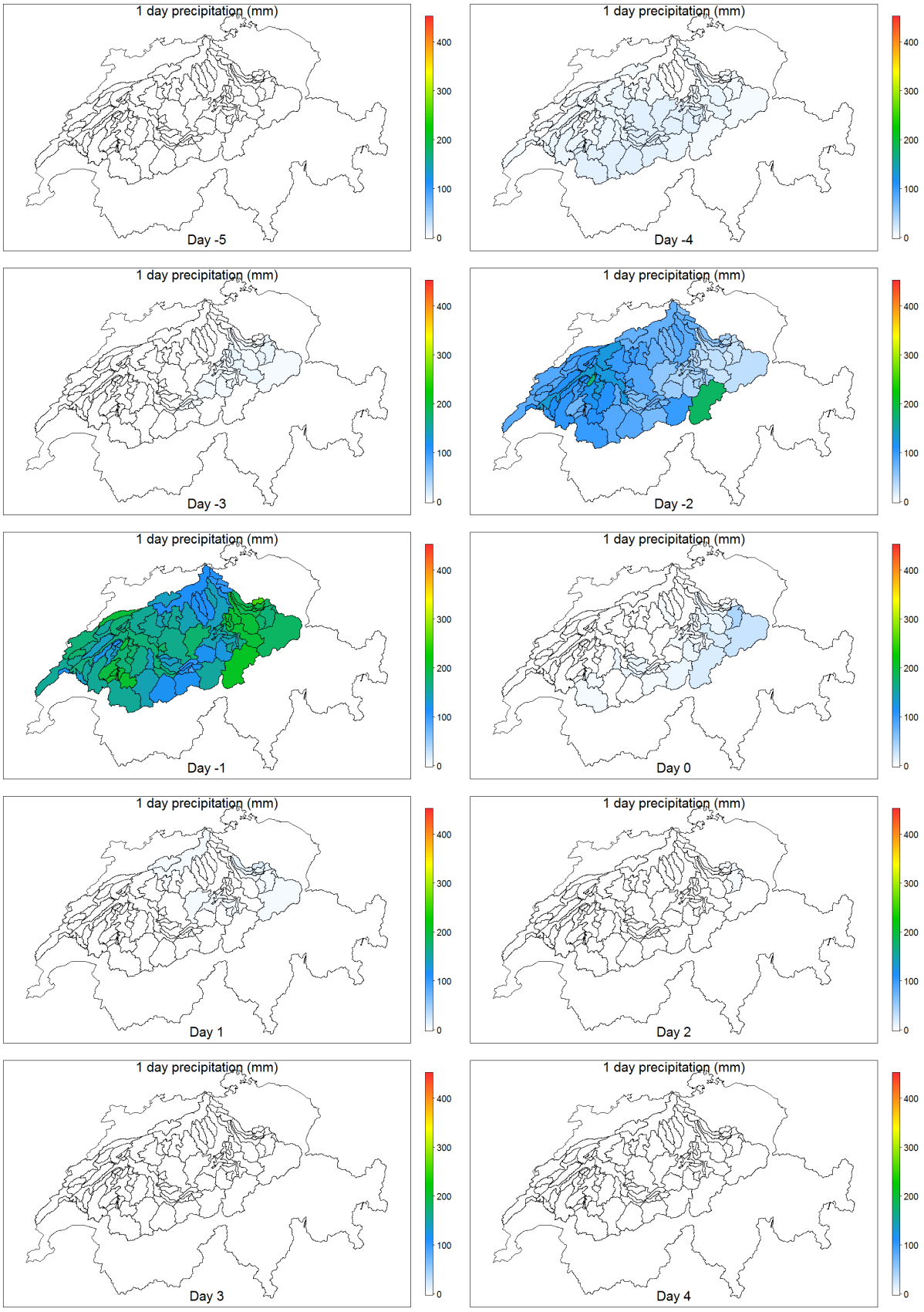


Figure A. 29 Space/time dynamics of precipitation during 10 days for the 3rd 3-day-precipitation event.

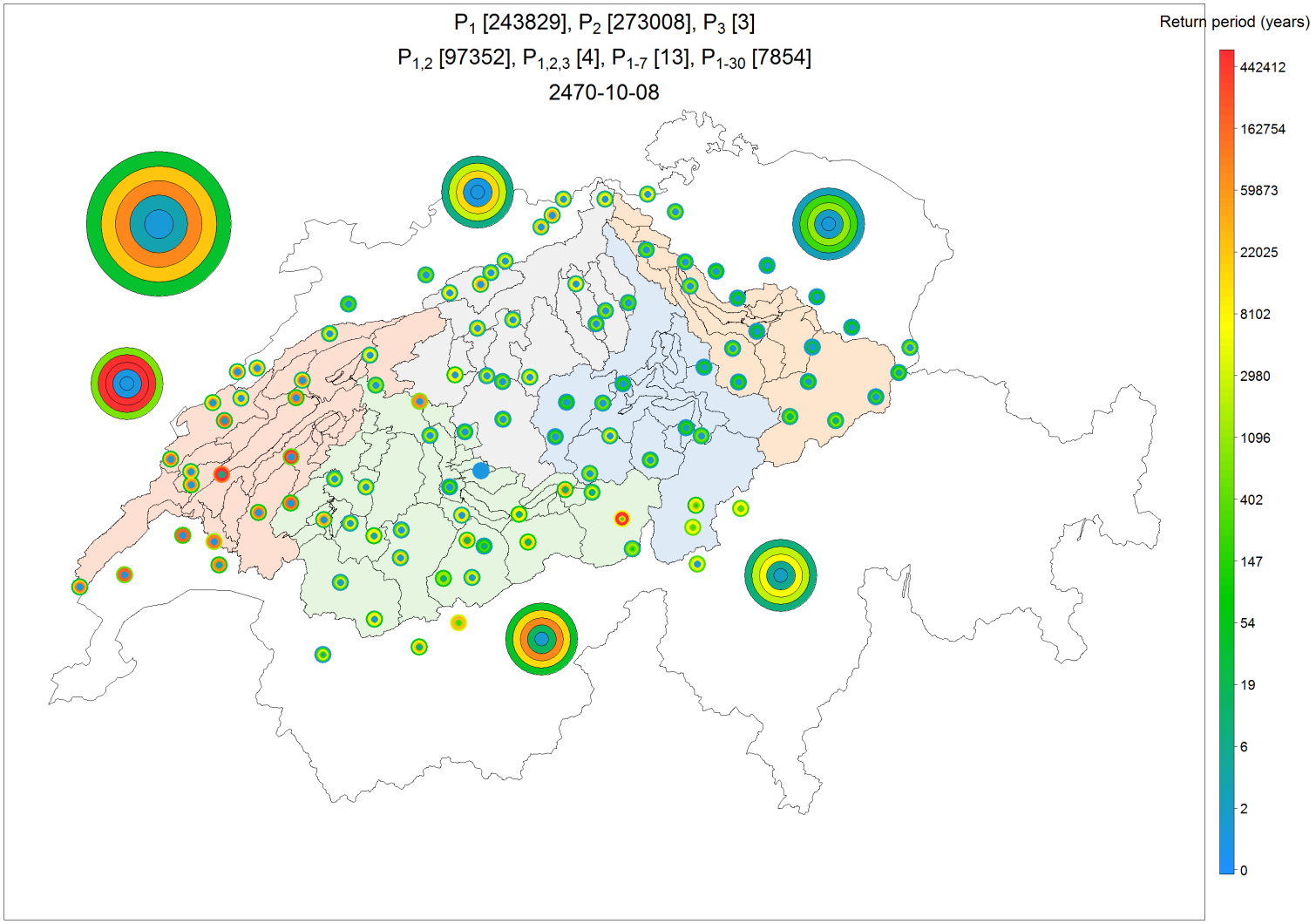


Figure A. 30 S Severity map representing the severity at several spatial scales and for several accumulation periods for the 4th 3-day precipitation event.

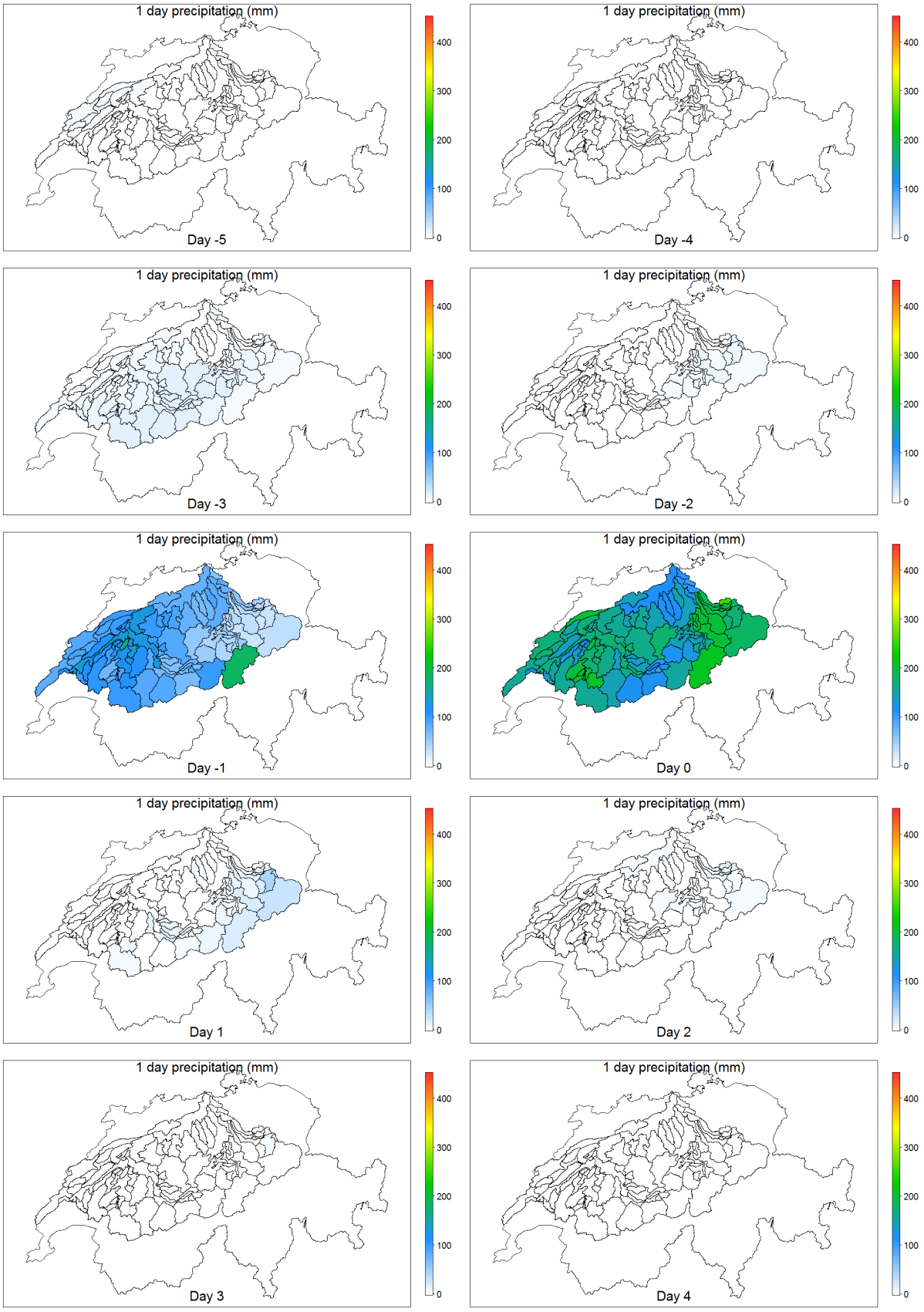


Figure A. 31 Space/time dynamics of precipitation during 10 days for the 4th 3-day-precipitation event.

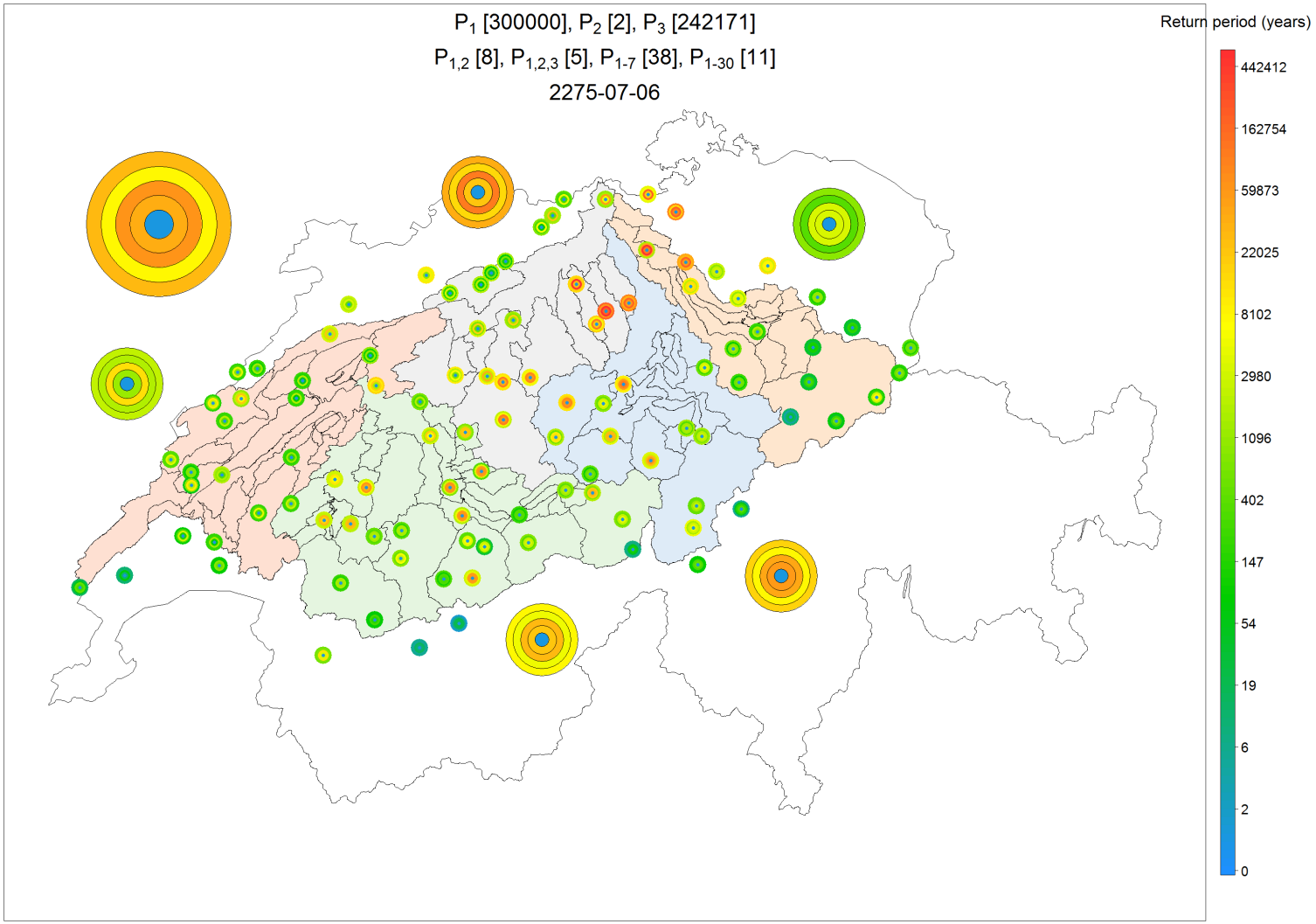


Figure A. 32 Severity map representing the severity at several spatial scales and for several accumulation periods for the 5th 3-day precipitation event.

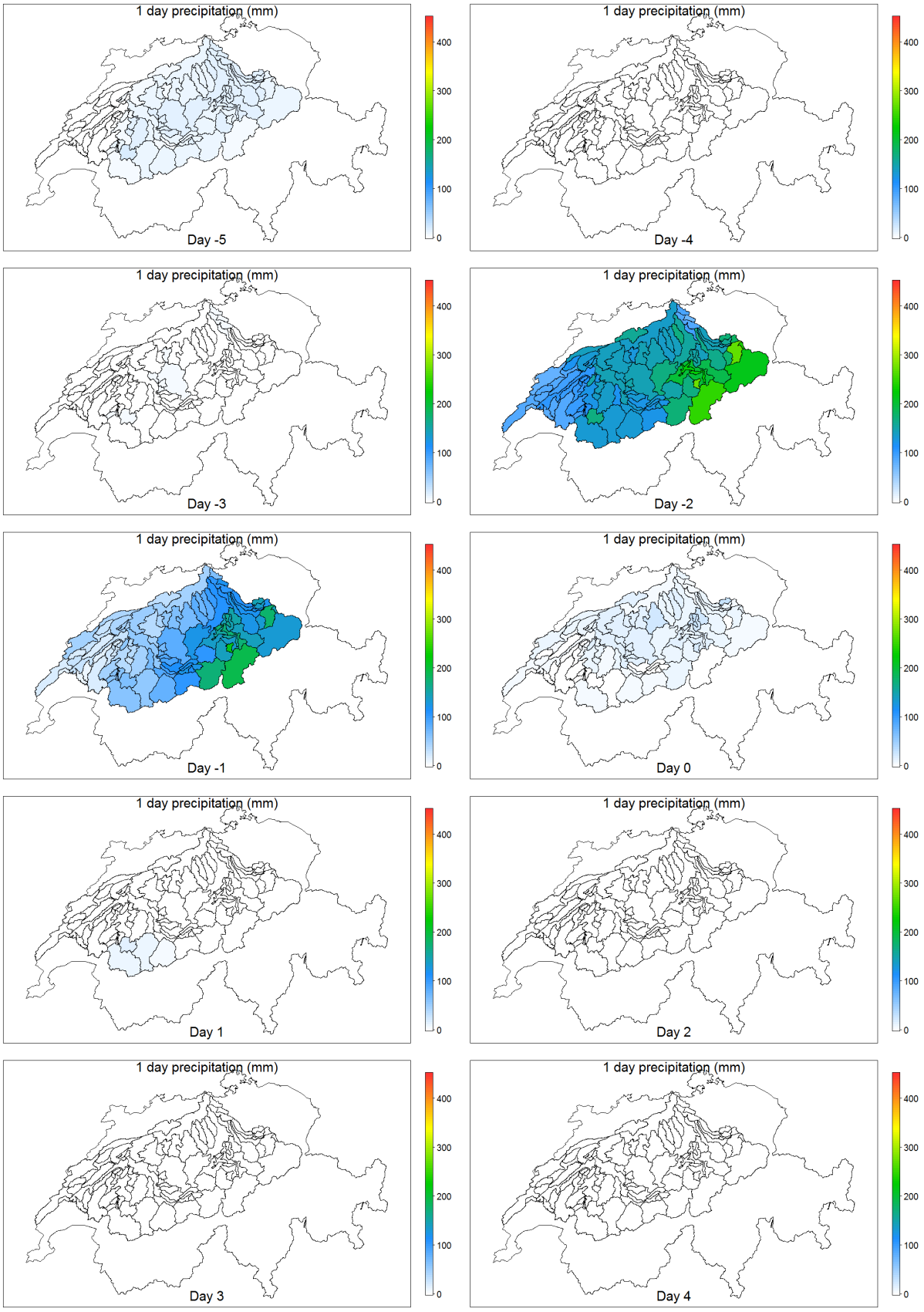


Figure A. 33 Space/time dynamics of precipitation during 10 days for the 5th 3-day-precipitation event.

10.5.3 Largest hydrological events with GWEX: Severity maps and space/time dynamics

The precipitation space-time patterns during the largest hydrological events are shown in the following figures: for each event first the severity map is shown, then the space/time dynamics of precipitation. The severity map represents the severity of this event at several spatial scales and for several accumulation periods; the space/time dynamics of precipitation for 10 days are composed of 10 maps ("Day 0" is the day of the discharge peak).

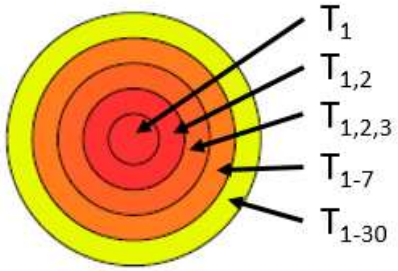
The title of the severity map shows the rank (in descending order) of the precipitation events at the whole catchment scale for several accumulation durations:

- H[]: rank of the hydrological event
- P_0[]: rank of the 1-day precipitation during the day of the hydrological peak (“Day 0”)
- P_1[]: rank of the 1-day precipitation during the day before the hydrological peak (“Day -1”)
- P_2[]: rank of the 1-day precipitation 2 days before the hydrological peak (“Day -2”)
- P_3[]: rank of the 1-day precipitation 3 days before the hydrological peak (“Day -3”)
- $P_{1,2}$[]: rank of the 2-day precipitation from “Day -1” to “Day -2”
- $P_{1,2,3}$[]: rank of the 3-day precipitation from “Day -1” to “Day -3”
- $P_{1\text{-}7}$[]: rank of the 7-day precipitation from “Day -1” to “Day -7”
- $P_{1\text{-}30}$[]: rank of the 30-day precipitation from “Day -1” to “Day -30”
- In addition, the fictive date of the hydrological peak is indicated on the third line

The severity of the simulated event is then represented with a set of severity-boards. Each severity-board gives the return period of precipitation simulated for this event when cumulated over several durations (all ending the day before the hydrological peak):

- 1 day (center of the severity-board)
- 2 days
- 3 days
- 7 days
- 30 days (edge of the severity-board)

For instance, if T_1, $T_{1,2}$, $T_{1,2,3}$, $T_{1\text{-}7}$, and $T_{1\text{-}30}$ are the return periods associated with the precipitation P_1, $P_{1,2}$, $P_{1,2,3}$, $P_{1\text{-}7}$, and $P_{1\text{-}30}$, respectively



The color scale for the severity-boards is given in the right-hand side of the figure. Severity-boards are given for precipitation simulated over different spatial scales:

- local precipitation at each of the 105 simulation stations (small circles)
- mean areal precipitation over each of the 5 main sub-catchments (medium circles)
- whole catchment (large circle)

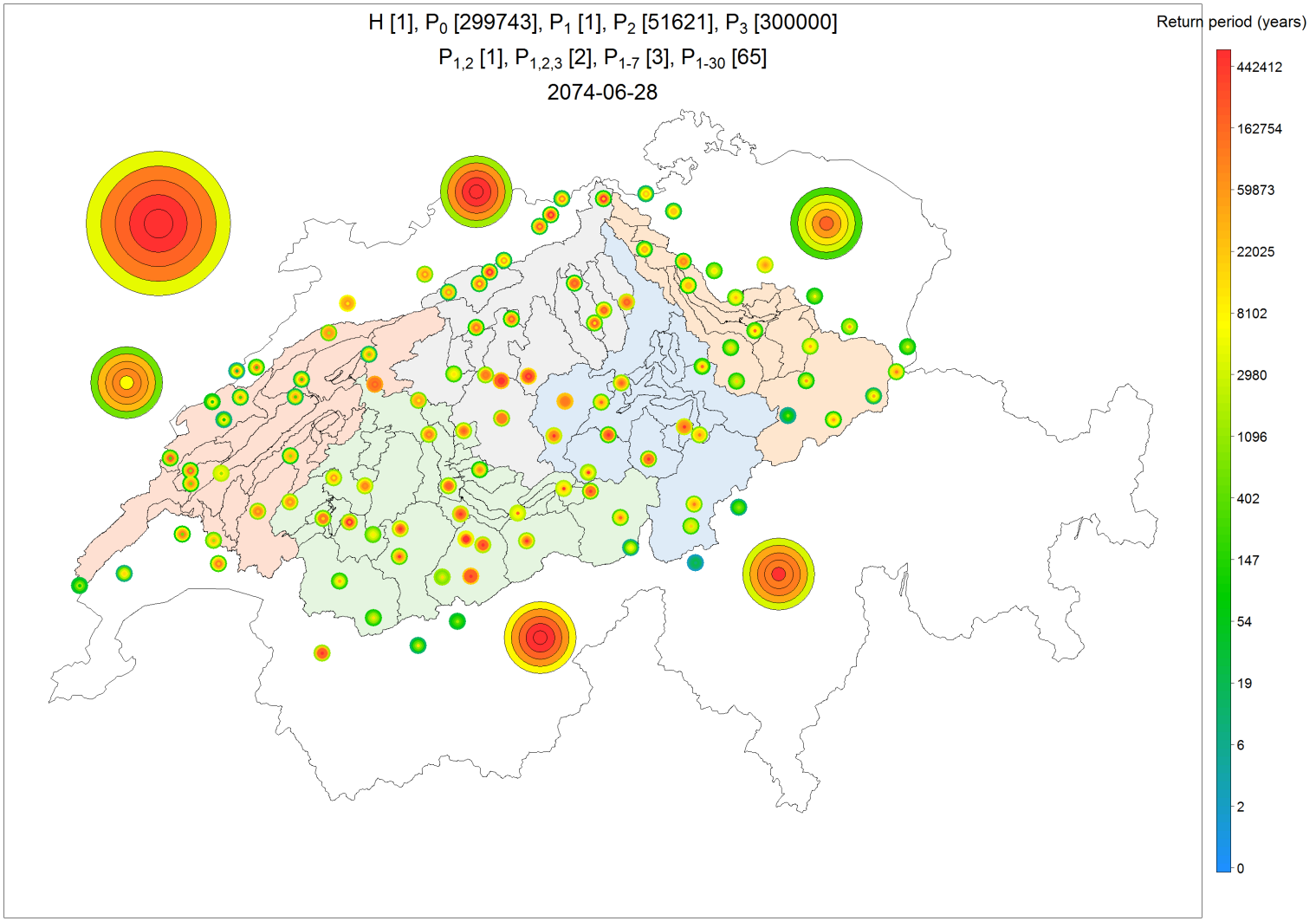


Figure A. 34 Severity map representing the severity at several spatial scales and for several accumulation periods for the 1st hydrological event

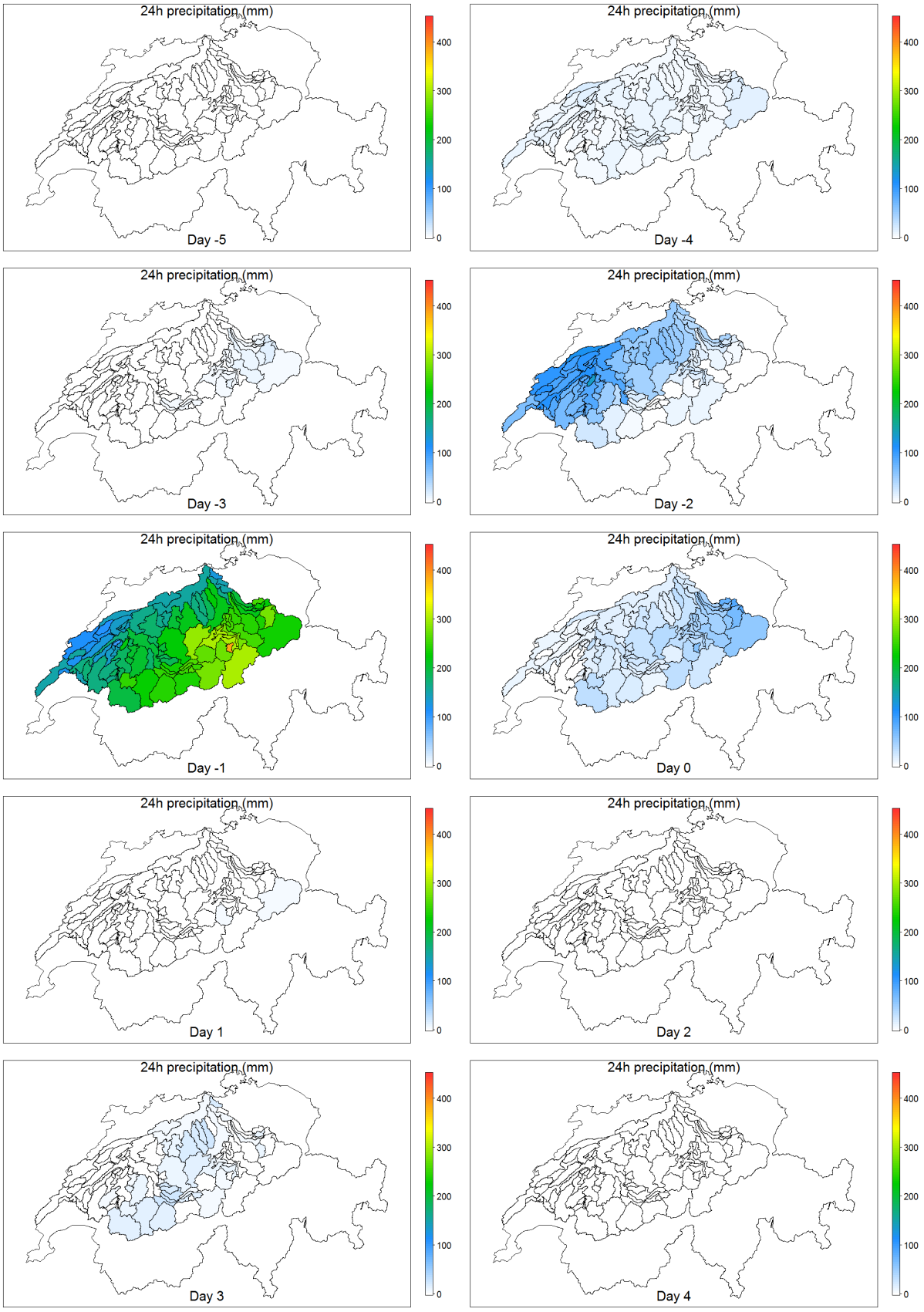


Figure A. 35 Space/time dynamics of precipitation during 10 days for the 1st hydrological event.

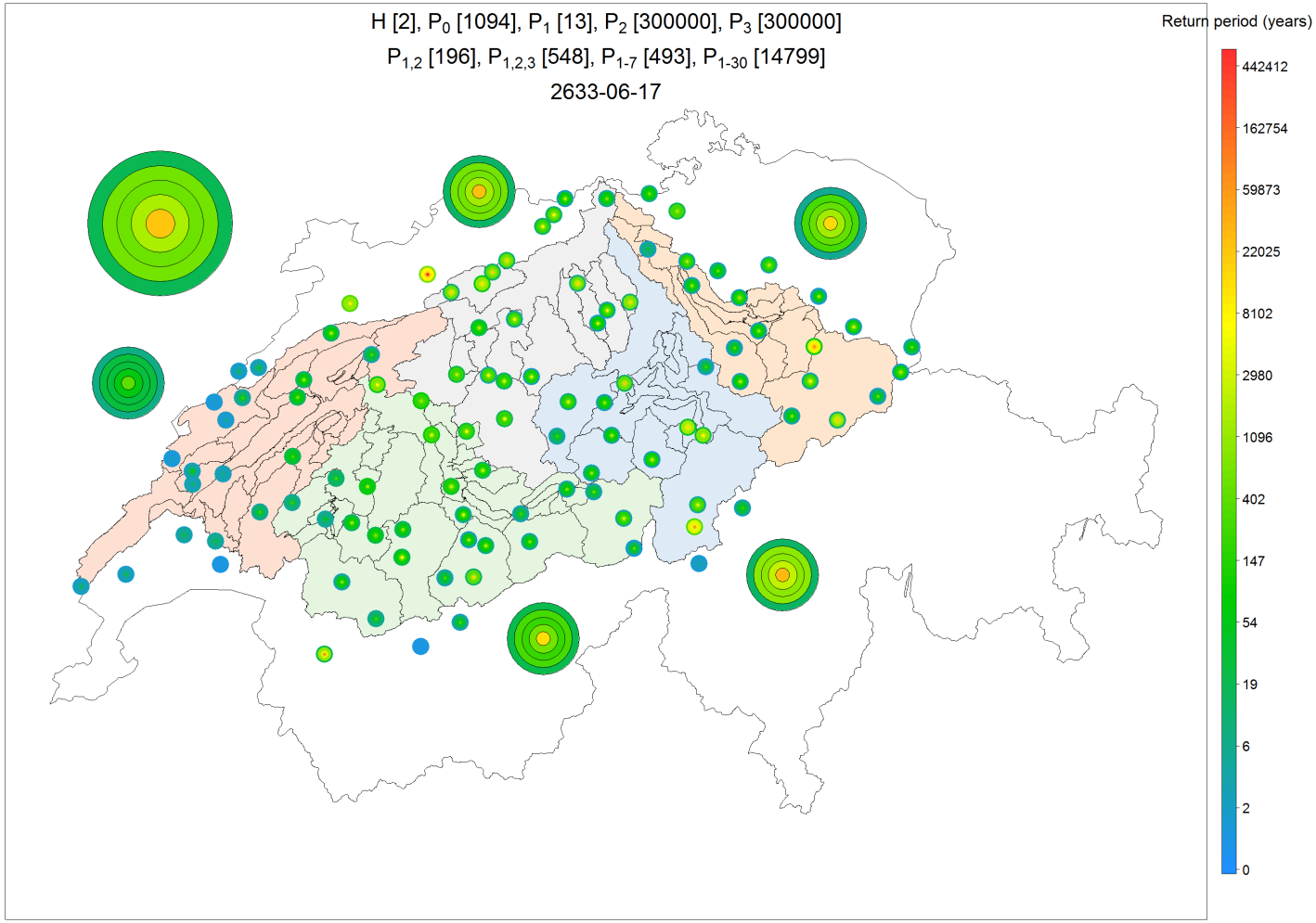


Figure A. 36 Severity map representing the severity at several spatial scales and for several accumulation periods for the 2nd hydrological event.

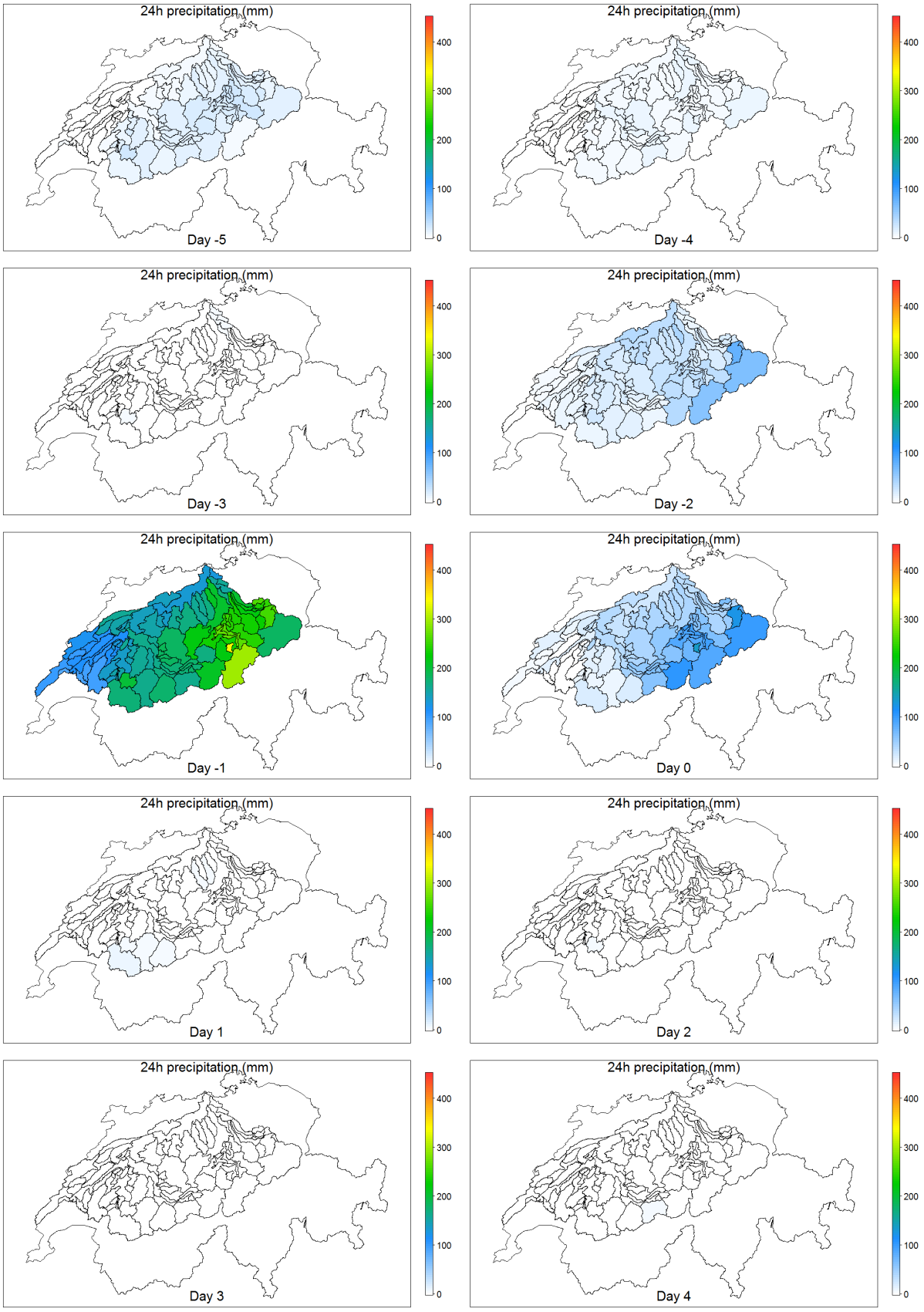


Figure A. 37 Space/time dynamics of precipitation during 10 days for the 2nd hydrological event.

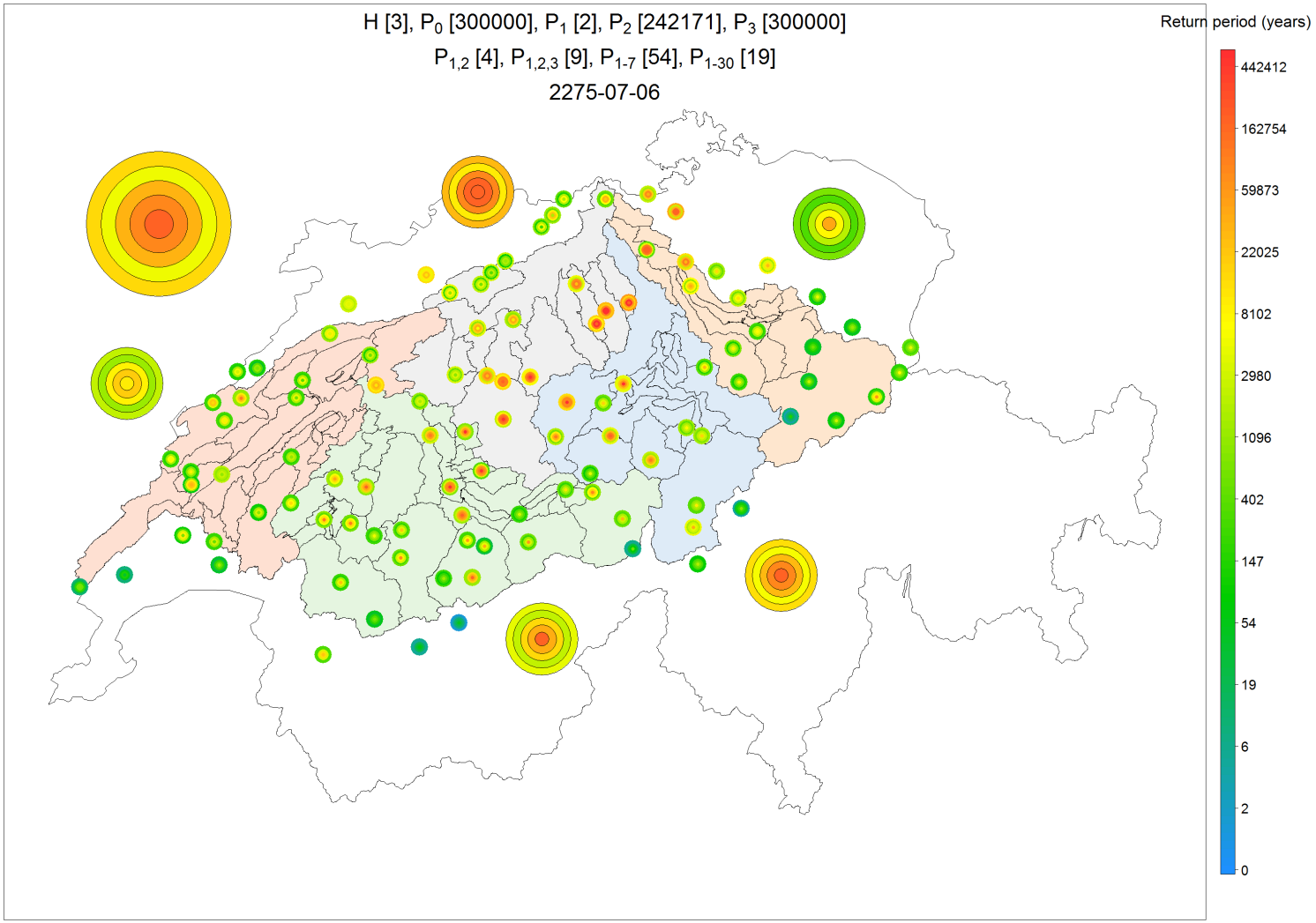


Figure A. 38 Severity map representing the severity at several spatial scales and for several accumulation periods for the 3rd hydrological event.

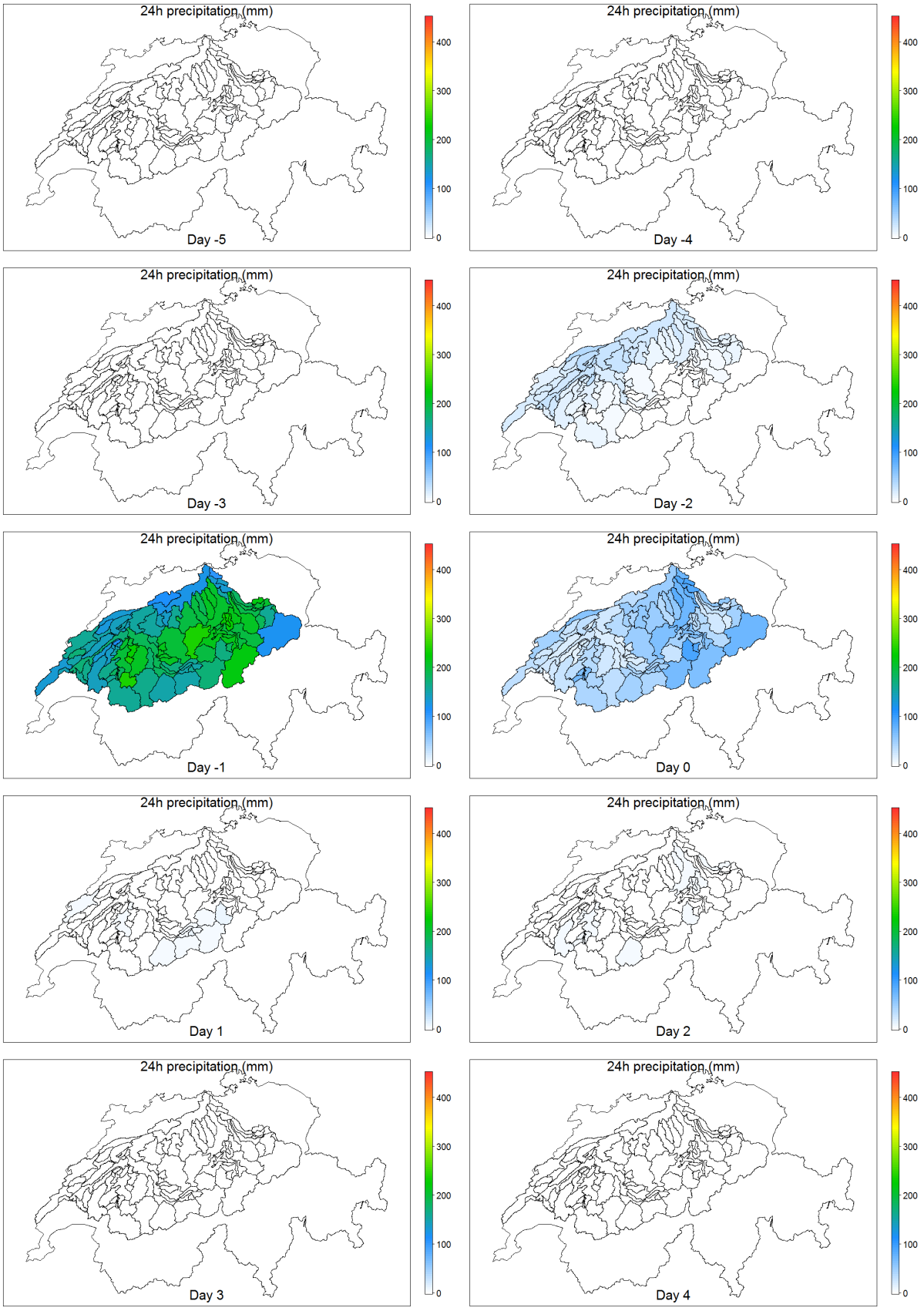


Figure A. 39 Space/time dynamics of precipitation during 10 days for the 3rd hydrological event.

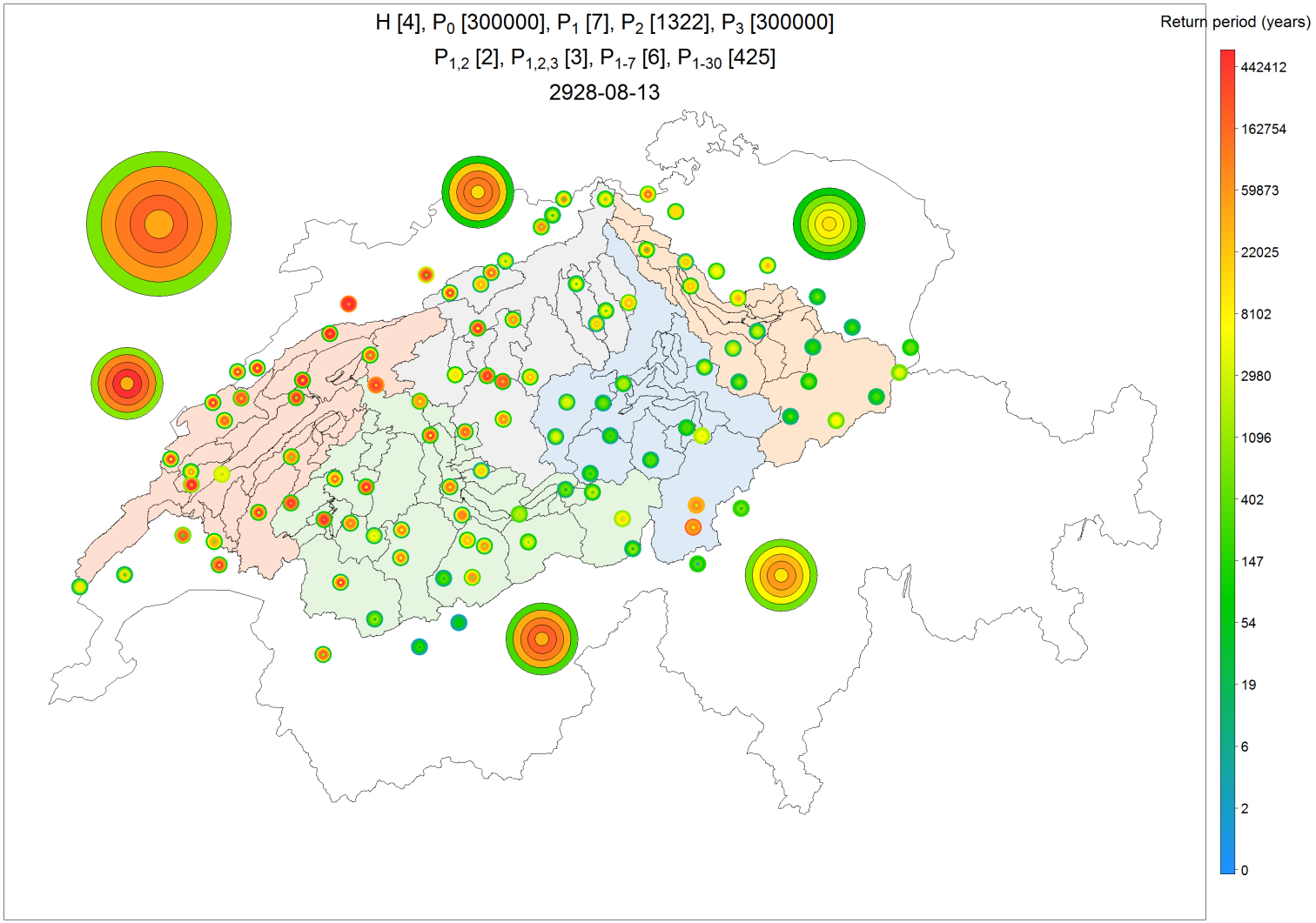


Figure A. 40 Severity map representing the severity at several spatial scales and for several accumulation periods for the 4th hydrological event.

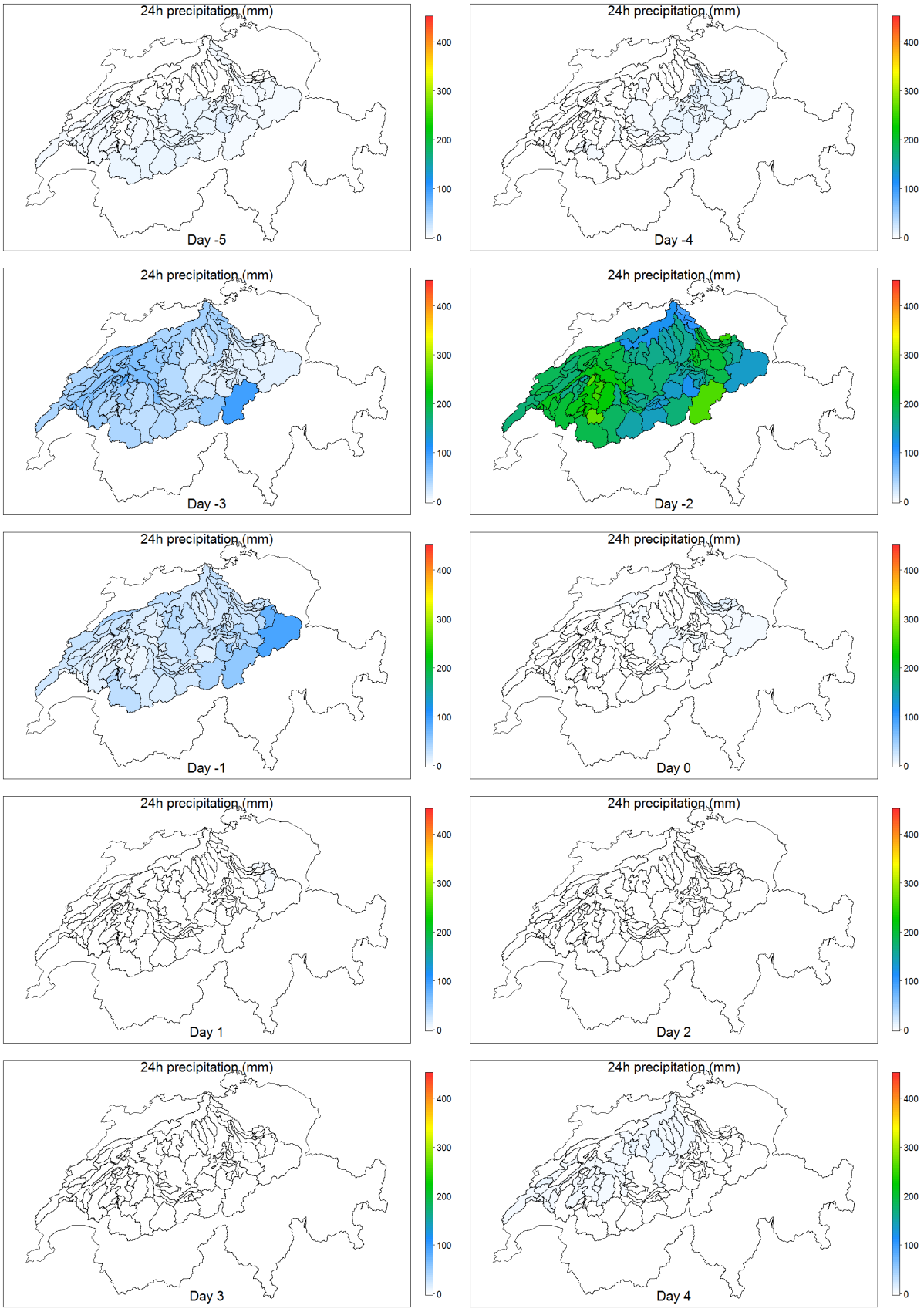


Figure A. 41 Space/time dynamics of precipitation during 10 days for the 4th hydrological event.

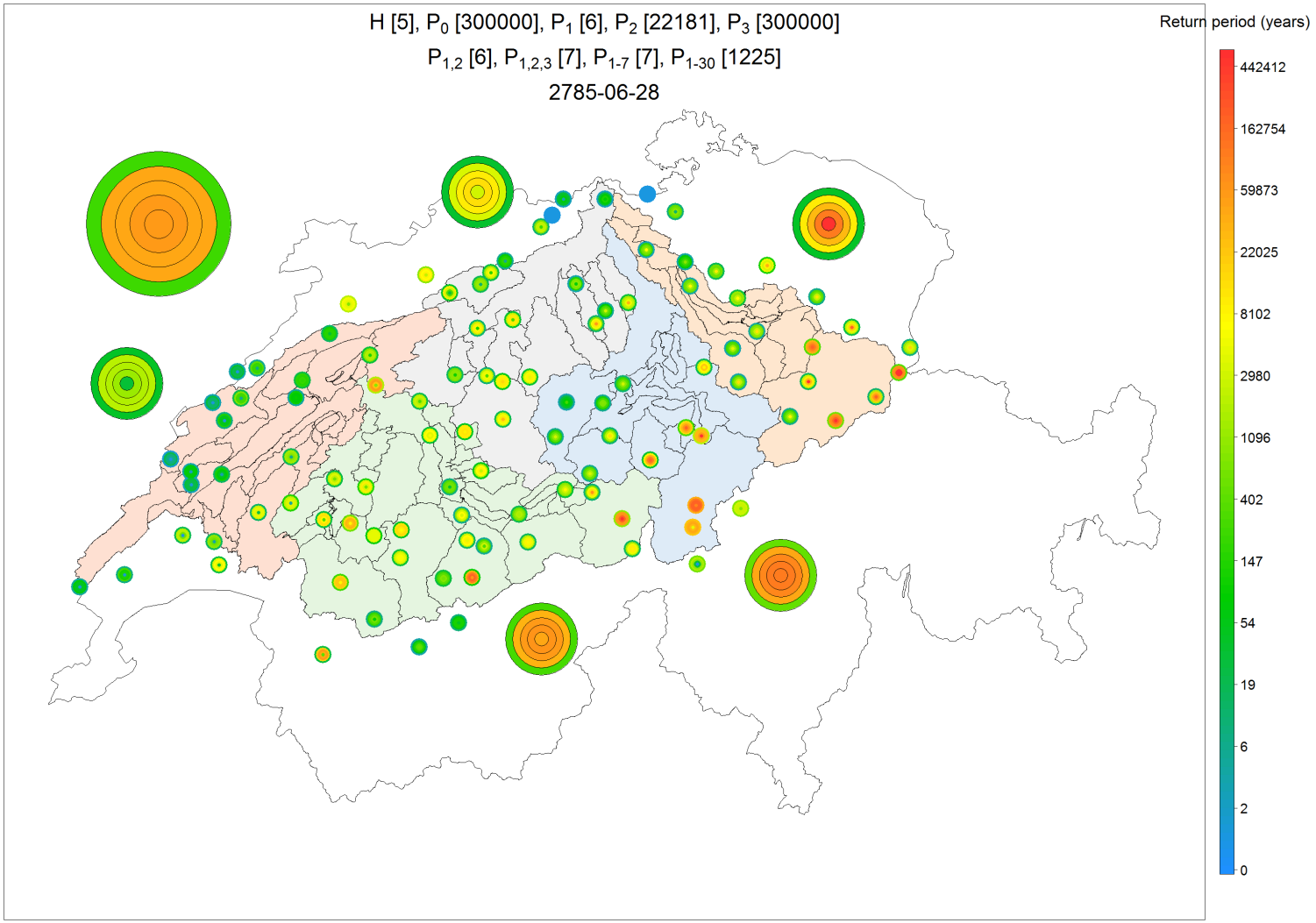


Figure A. 42 Severity map representing the severity at several spatial scales and for several accumulation periods for the 5th hydrological event

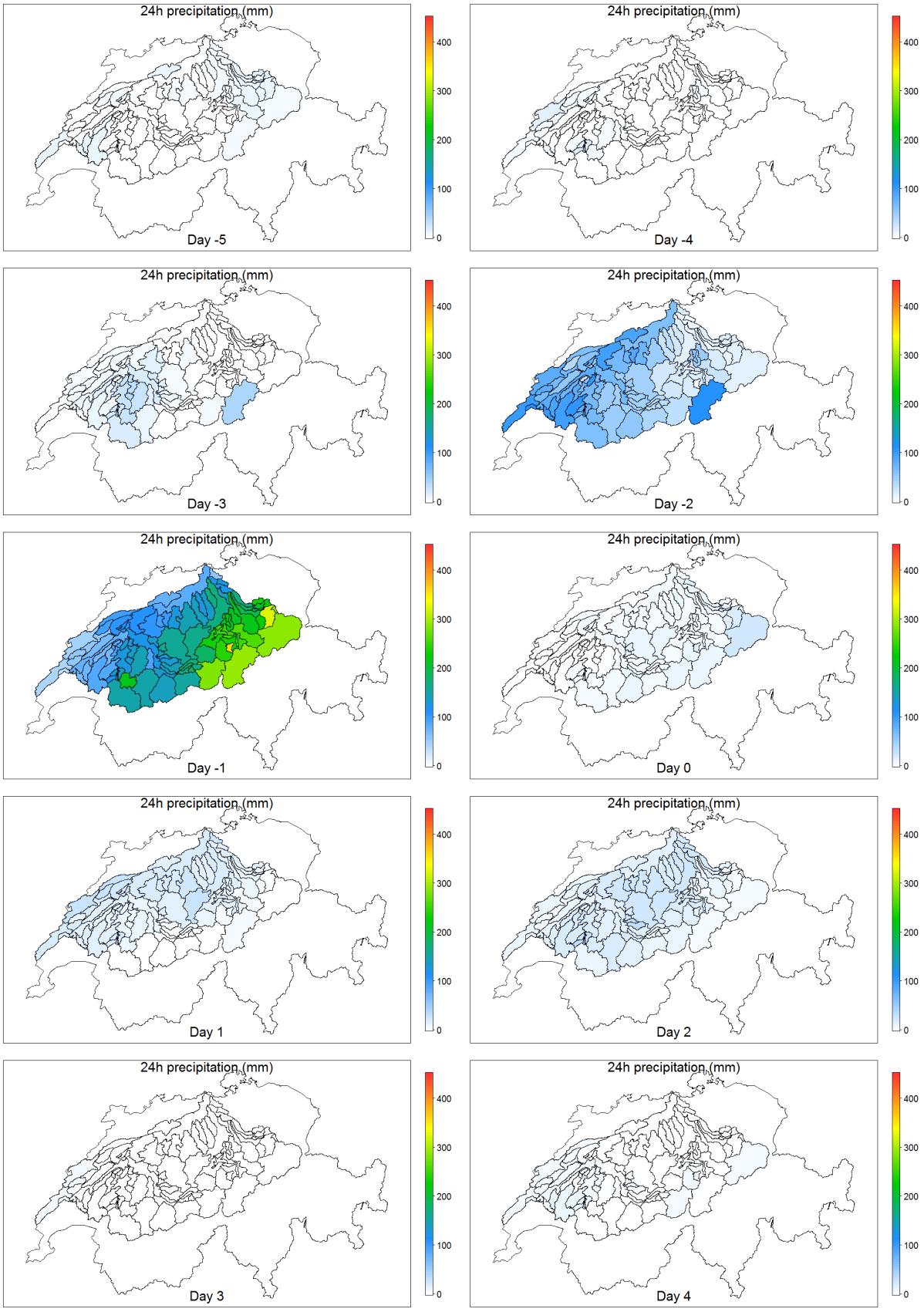


Figure A. 43 Space/time dynamics of precipitation during 10 days for the 5th hydrological event.

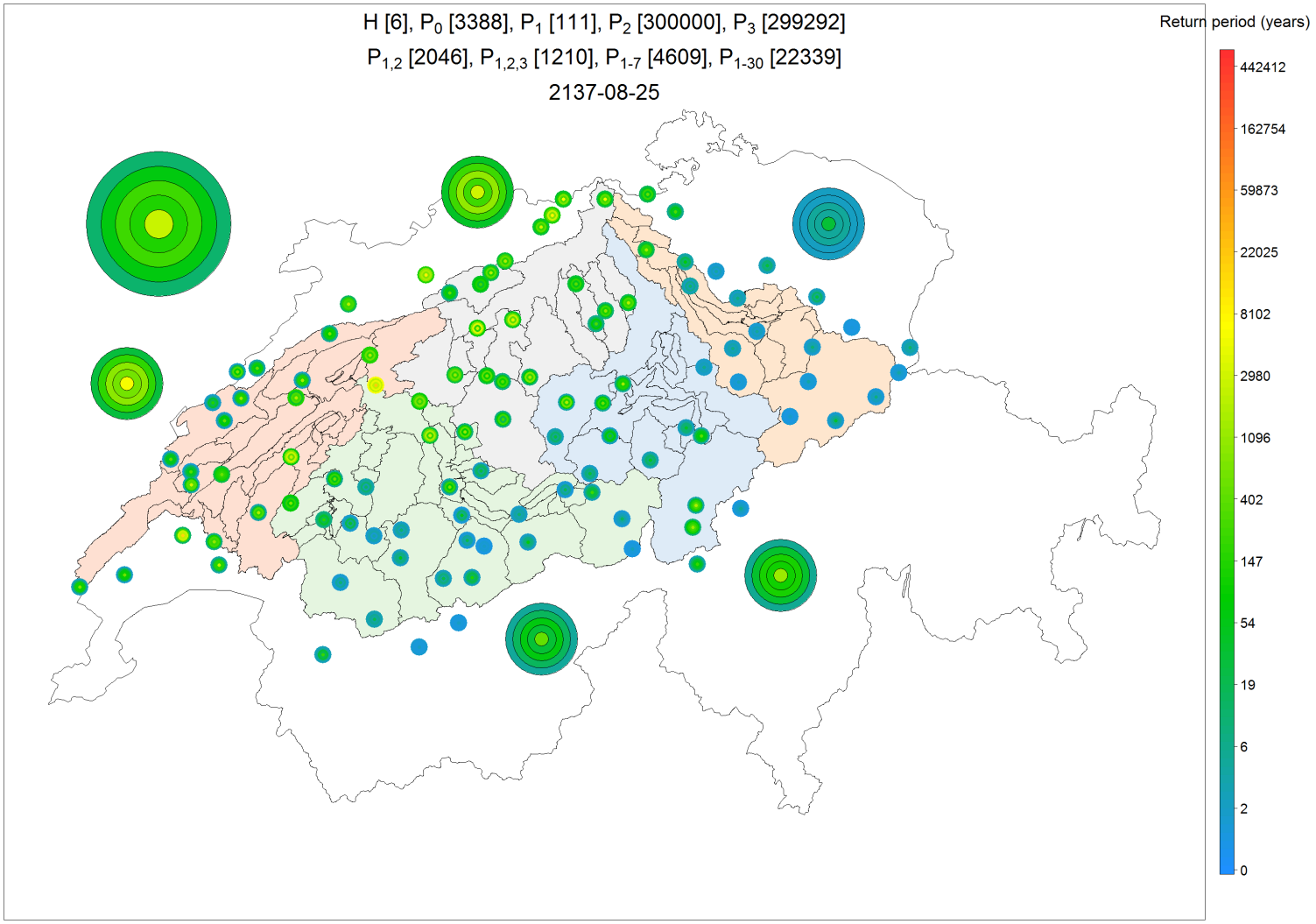


Figure A. 44 Severity map representing the severity at several spatial scales and for several accumulation periods for the 6th hydrological event.

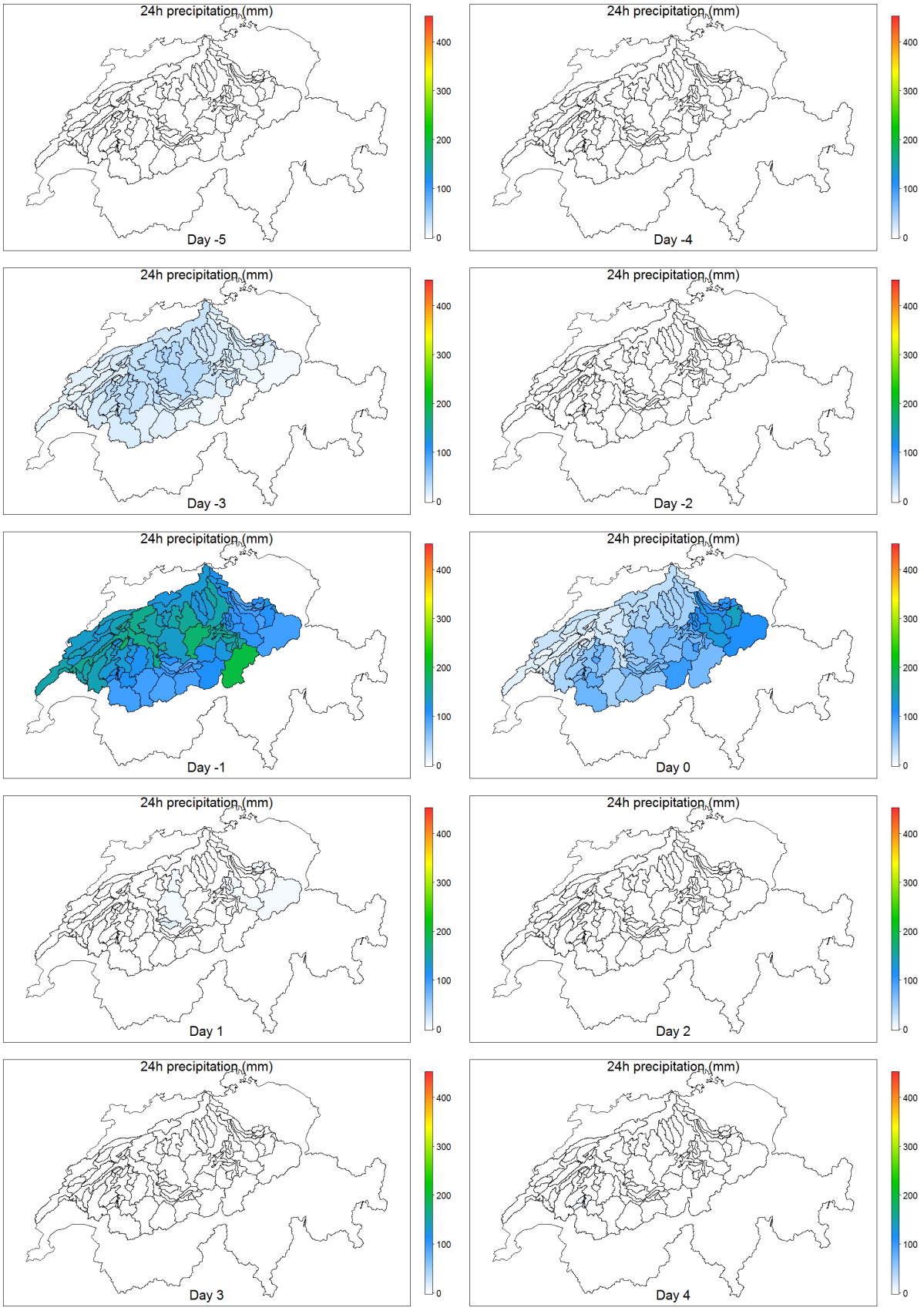


Figure A. 45 Space/time dynamics of precipitation during 10 days for the 6th hydrological event.

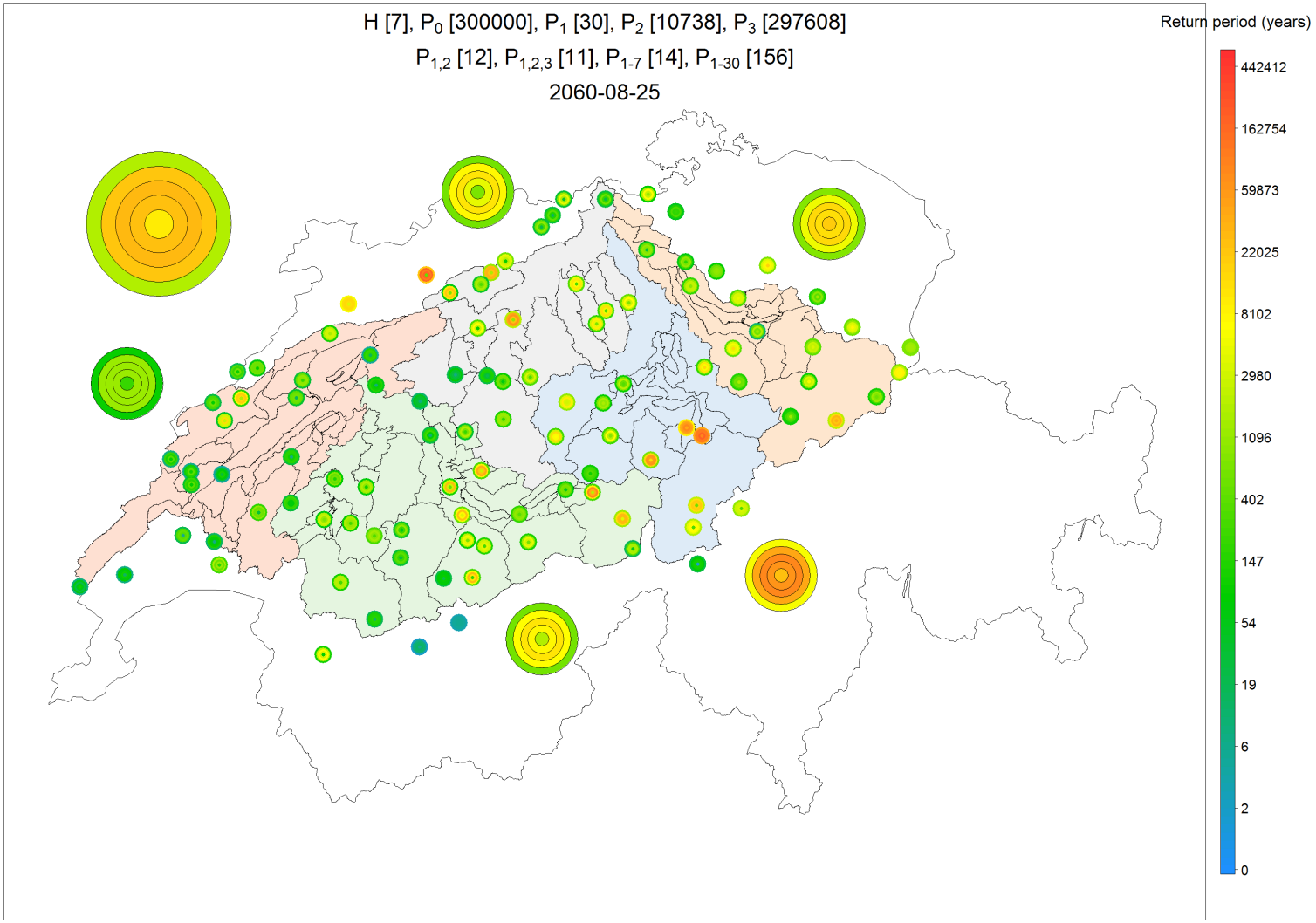


Figure A. 46 Severity map representing the severity at several spatial scales and for several accumulation periods for the 7th hydrological event.

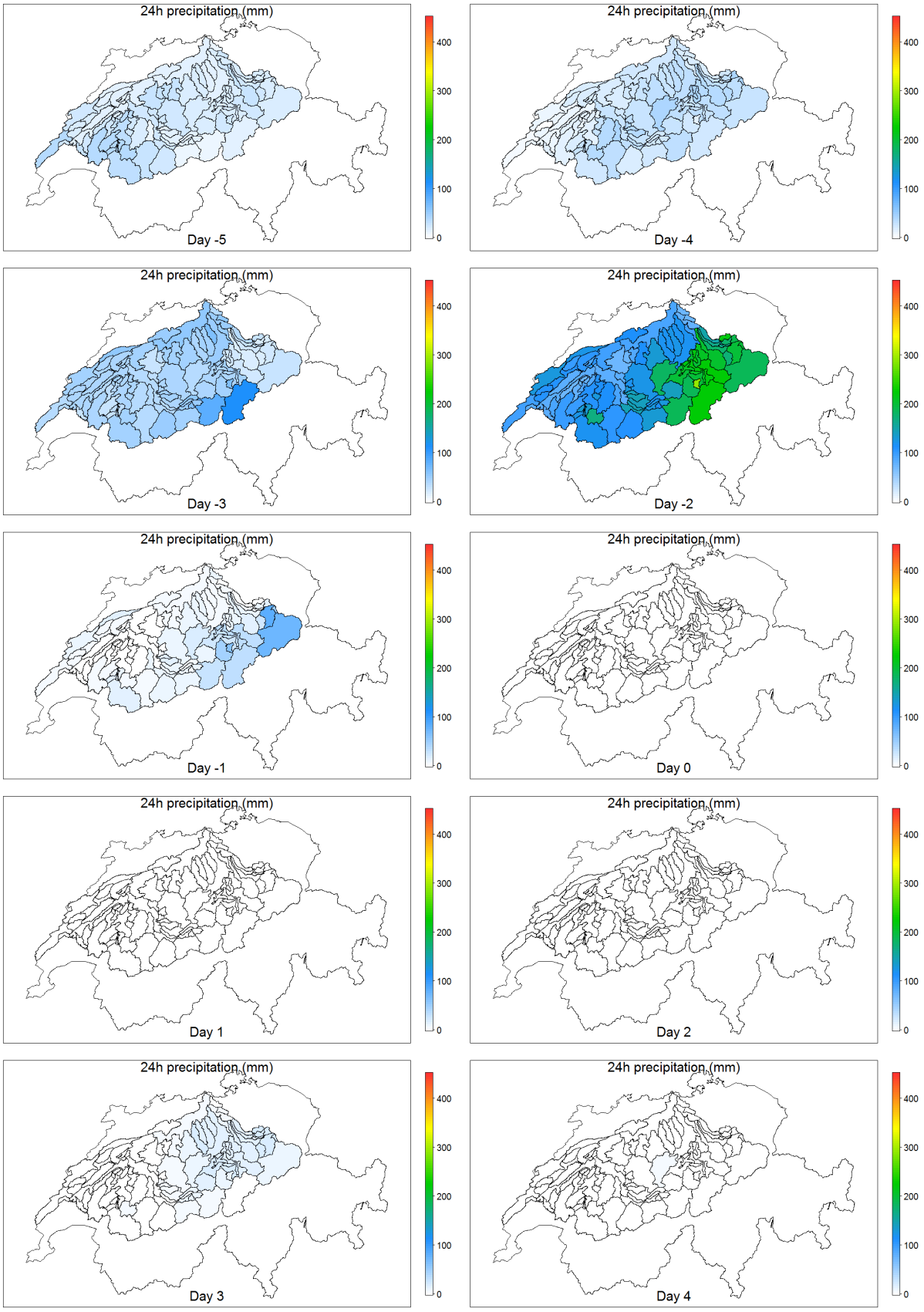


Figure A. 47 Space/time dynamics of precipitation during 10 days for the 7th hydrological event.

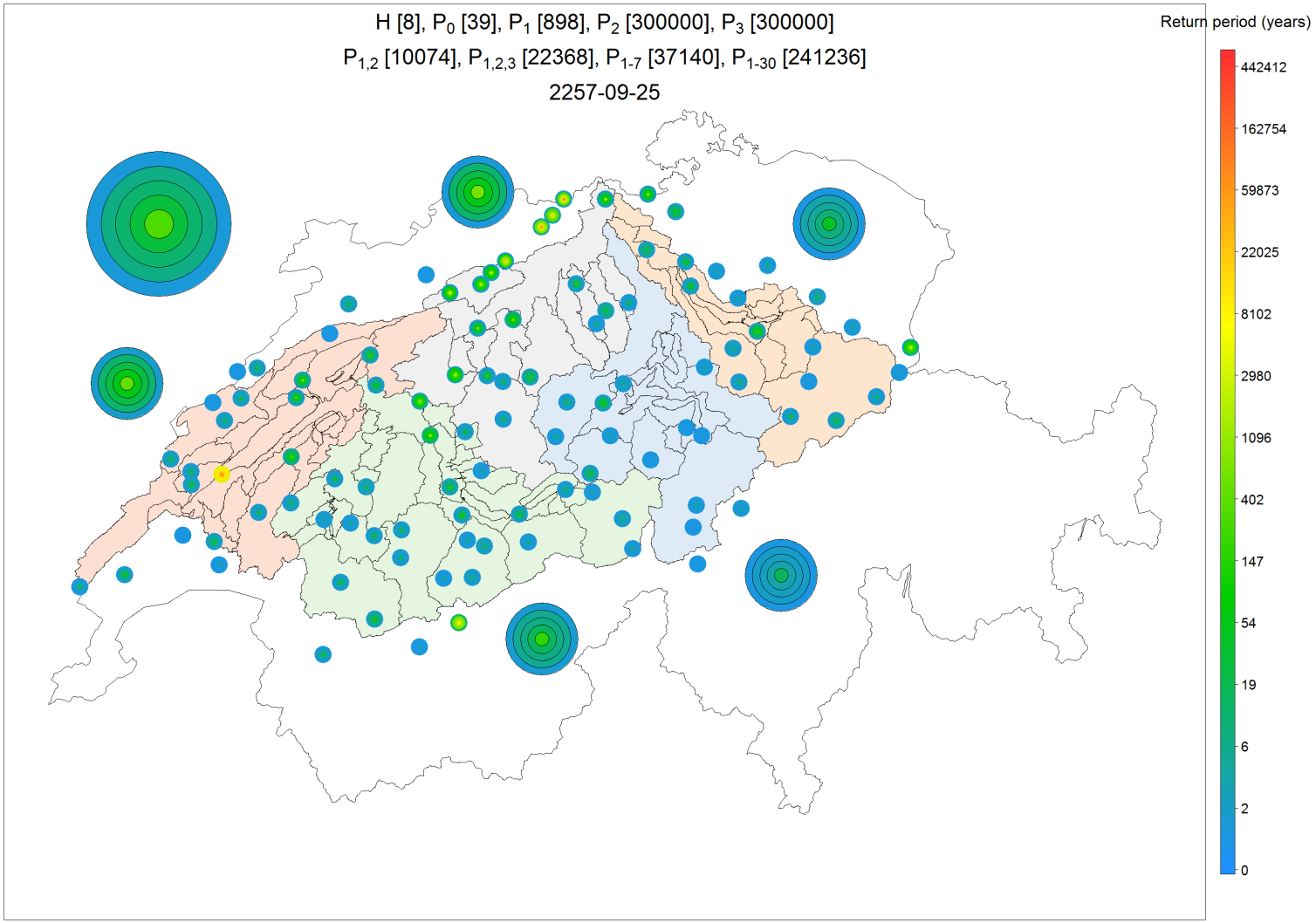


Figure A. 48 Severity map representing the severity at several spatial scales and for several accumulation periods for the 8th hydrological event.

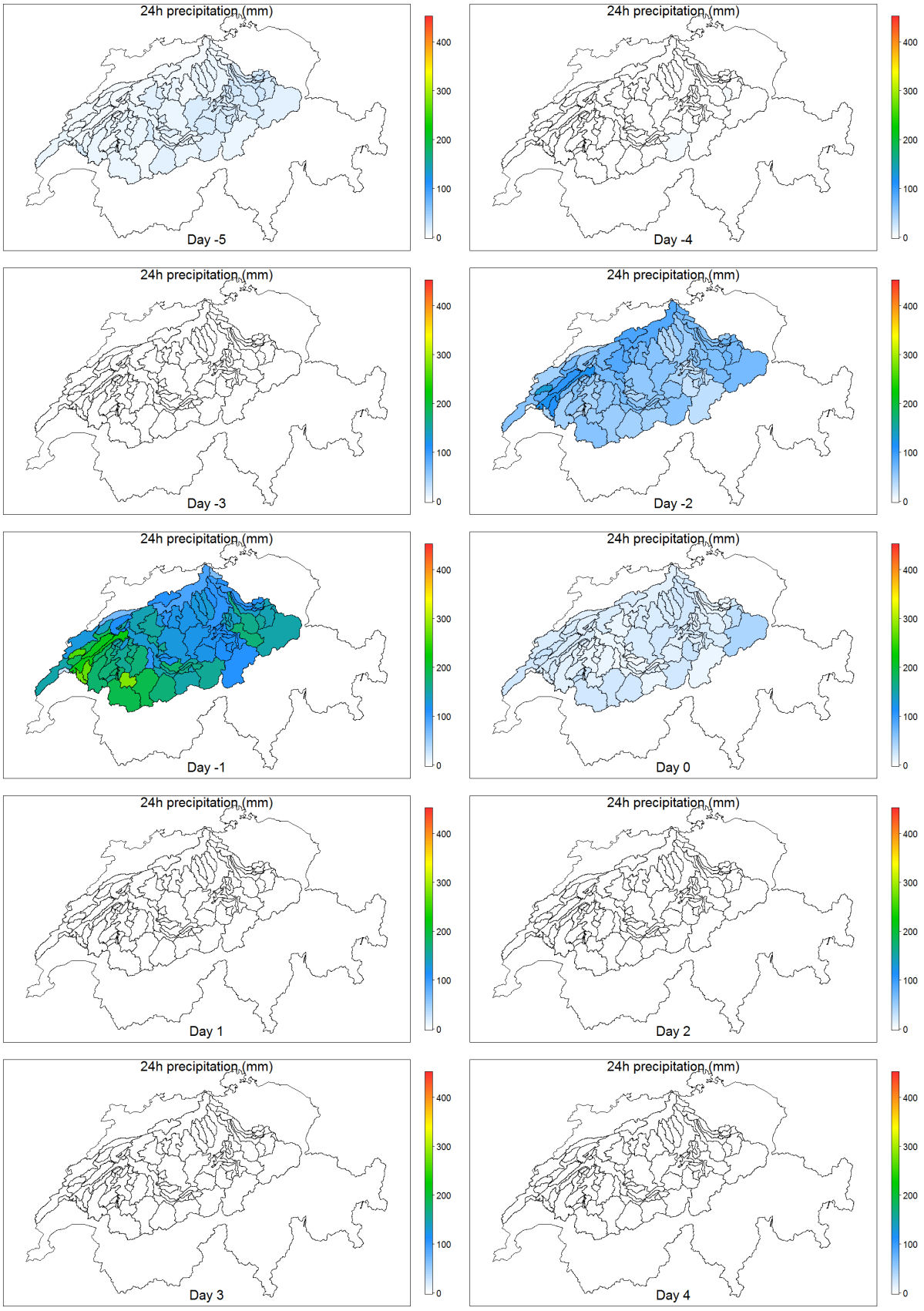


Figure A. 49 Space/time dynamics of precipitation during 10 days for the 8th hydrological event.

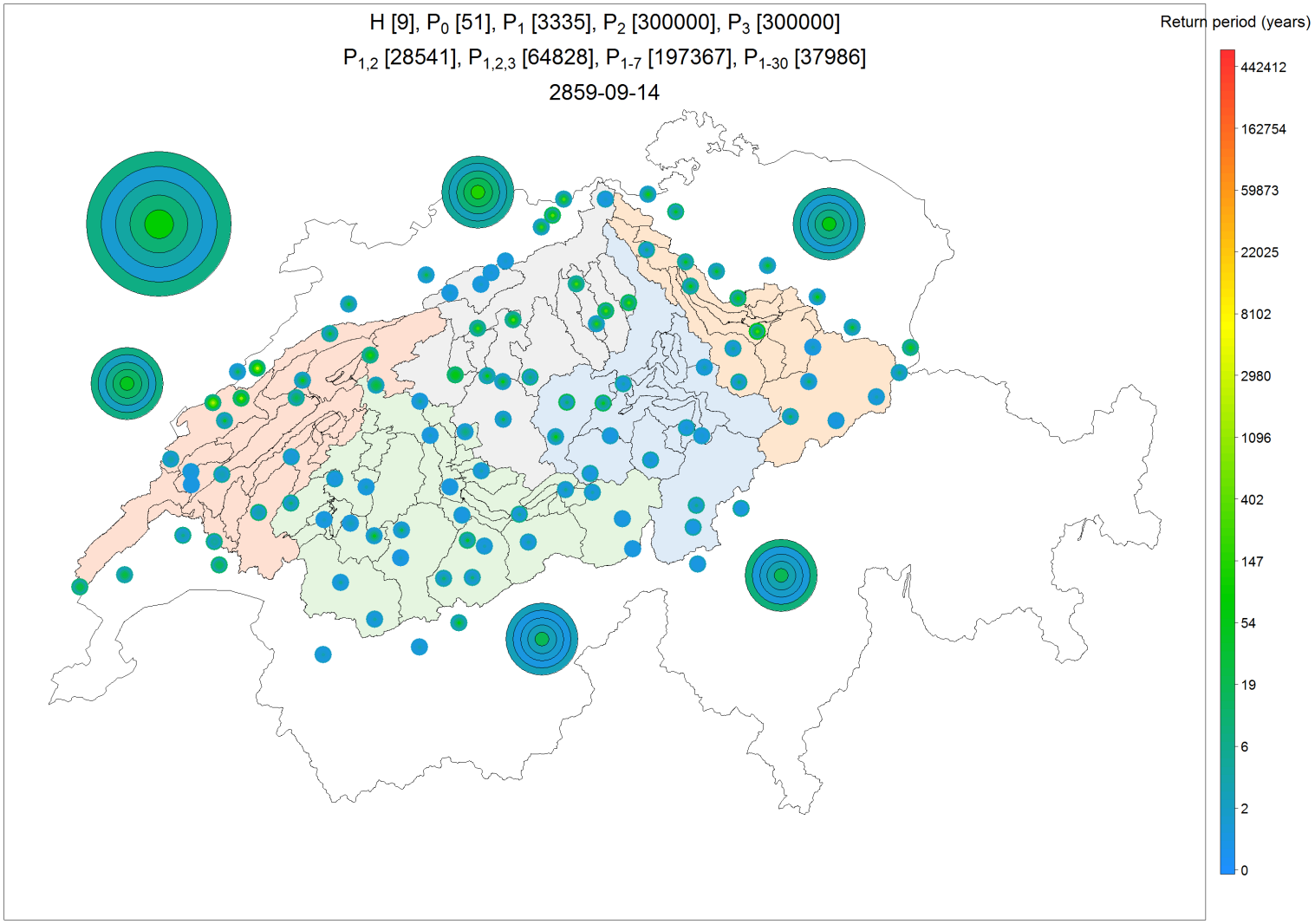


Figure A. 50 Severity map representing the severity at several spatial scales and for several accumulation periods for the 9th hydrological event.

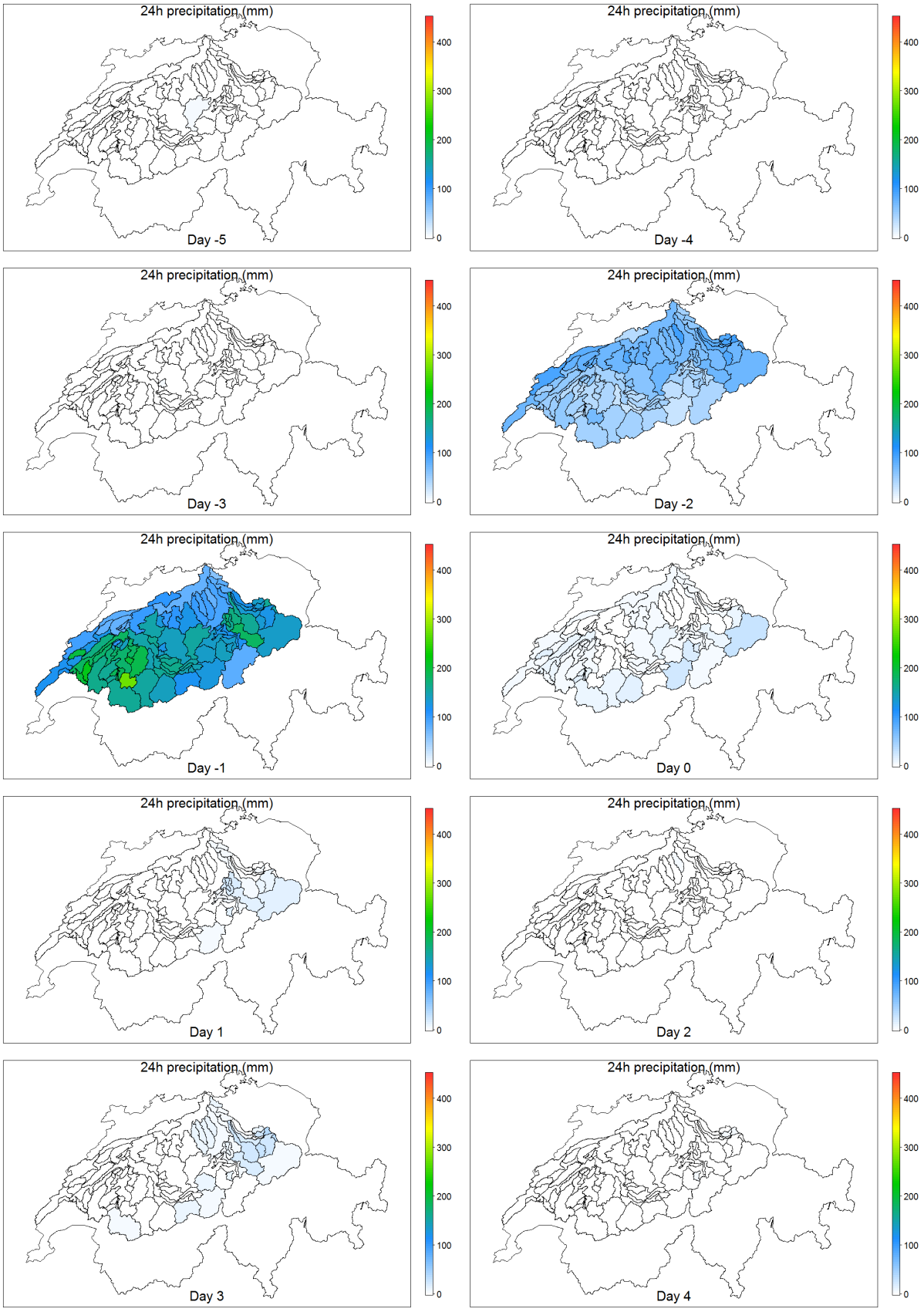


Figure A. 51 Space/time dynamics of precipitation during 10 days for the 9th hydrological event.

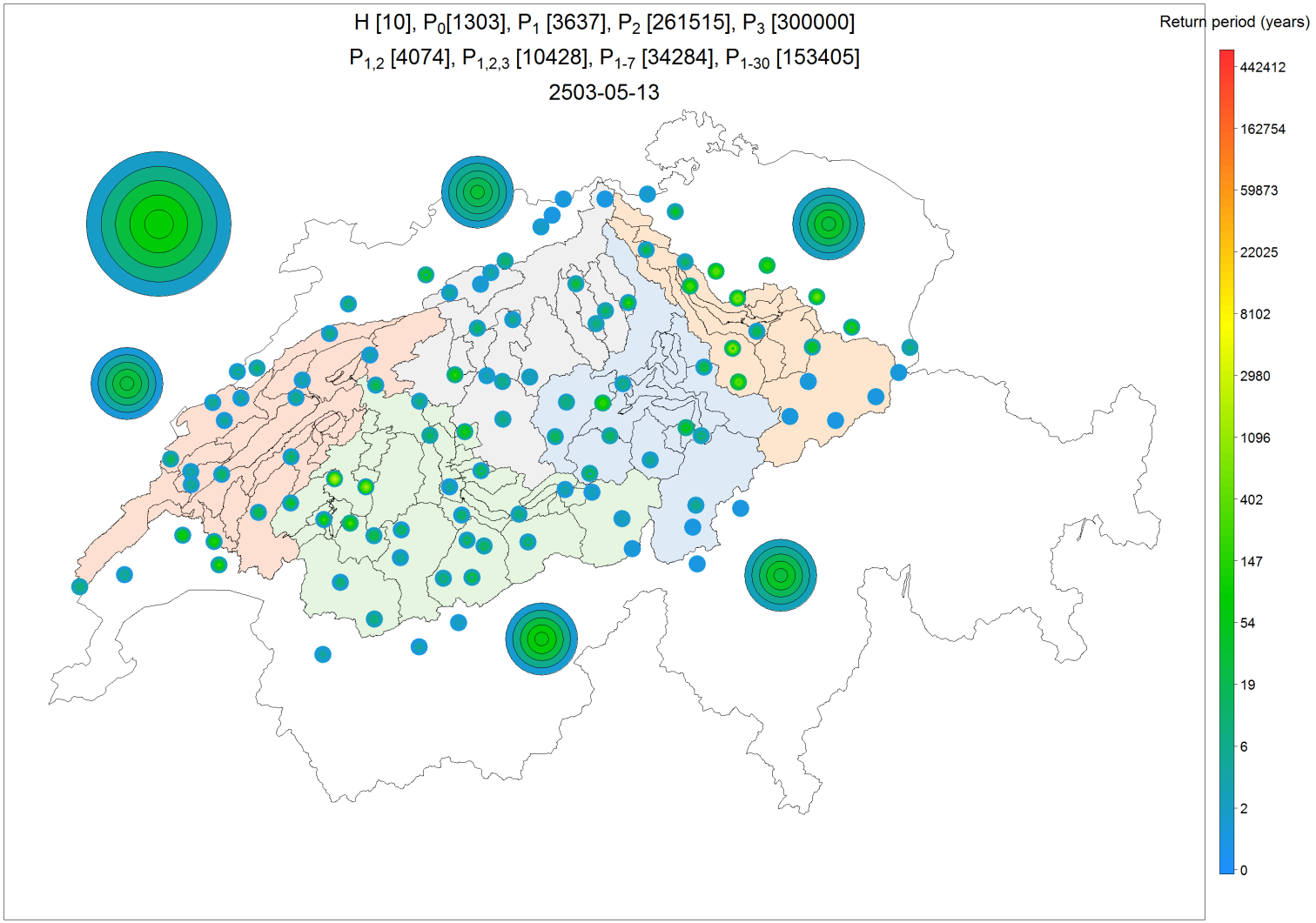


Figure A. 52 Severity map representing the severity at several spatial scales and for several accumulation periods for the 10th hydrological event.

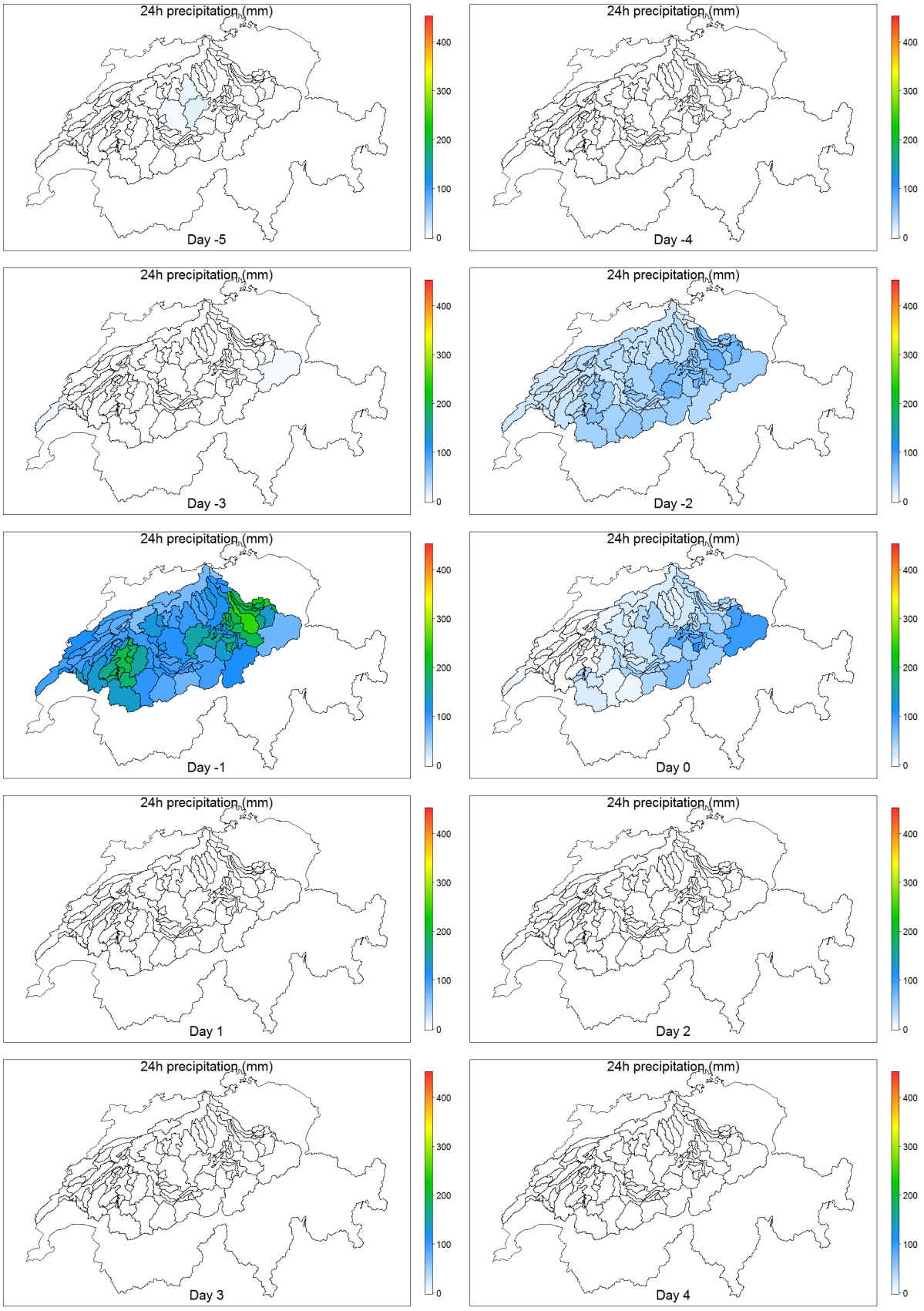


Figure A. 53 Space/time dynamics of precipitation during 10 days for the 10th hydrological event.

10.5.4 Largest precipitation events with SCAMP: Severity maps and space/time dynamics

The precipitation space-time patterns during the largest precipitation events with SCAMP are shown in the following figures: for each event first the severity map is shown, then the space/time dynamics of precipitation. The severity map represents the severity of this event at several spatial scales and for several accumulation periods; the space/time dynamics of precipitation for 10 days are composed of 10 maps ("Day 0" is the day of the discharge peak).

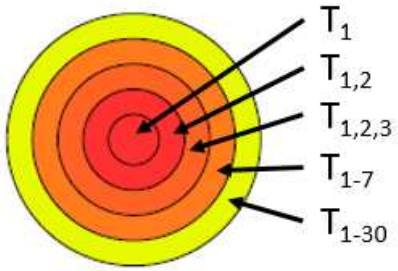
The title of the severity map shows the rank (in descending order) of the precipitation events at the whole catchment scale for several accumulation durations (all ending the day of the precipitation peak):

- P_1[]: rank of the 1-day precipitation during the day of the precipitation peak (“Day 0”)
- P_2[]: rank of the 1-day precipitation during the day before the precipitation peak
- P_3[]: rank of the 1-day precipitation 2 days before the precipitation peak
- $P_{1,2}$[]: rank of the 2-day precipitation
- $P_{1,2,3}$[]: rank of the 3-day precipitation
- $P_{1\text{-}7}$[]: rank of the 7-day precipitation
- $P_{1\text{-}30}$[]: rank of the 30-day precipitation
- In addition, the fictive date of the precipitation peak is indicated on the third line

The severity of the simulated event is then represented with a set of severity-boards. Each severity-board gives the return period of precipitation simulated for this event when cumulated over several durations (all ending the day of the precipitation peak):

- 1 day (center of the severity-board)
- 2 days
- 3 days
- 7 days
- 30 days (edge of the severity-board)

For instance, if T_1, $T_{1,2}$, $T_{1,2,3}$, $T_{1\text{-}7}$, and $T_{1\text{-}30}$ are the return periods associated with the precipitation P_1, $P_{1,2}$, $P_{1,2,3}$, $P_{1\text{-}7}$, and $P_{1\text{-}30}$, respectively



The color scale for the severity-boards is given in the right-hand side of the figure. Severity-boards are given for precipitation simulated over different spatial scales:

- local precipitation at each of the 89 simulation sub-catchments (small circles)
- mean areal precipitation over each of the 5 main sub-catchments (medium circles)
- whole catchment (large circle)

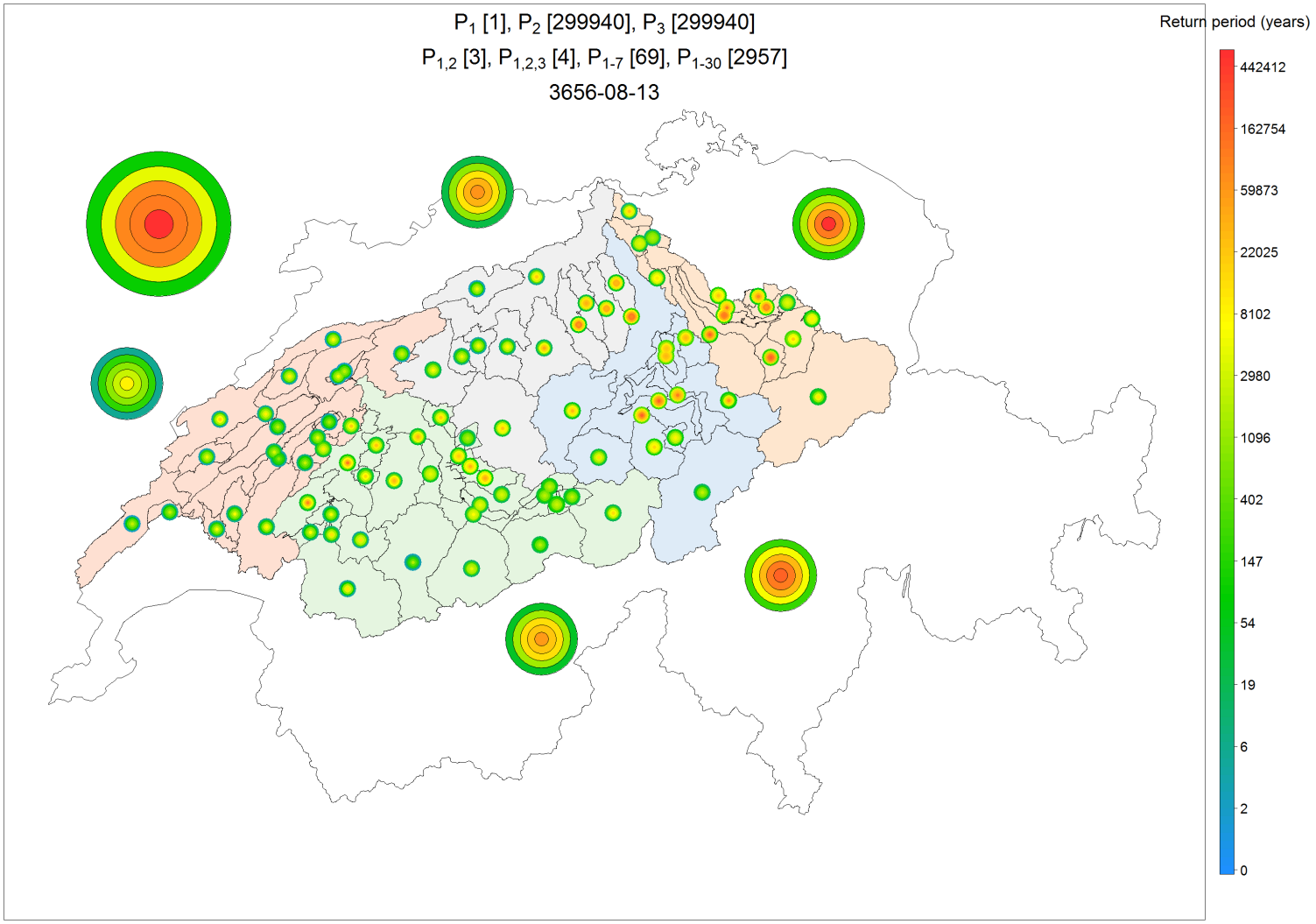


Figure A. 54 Severity map representing the severity at several spatial scales and for several accumulation periods for the 1st 1-day-precipitation event.

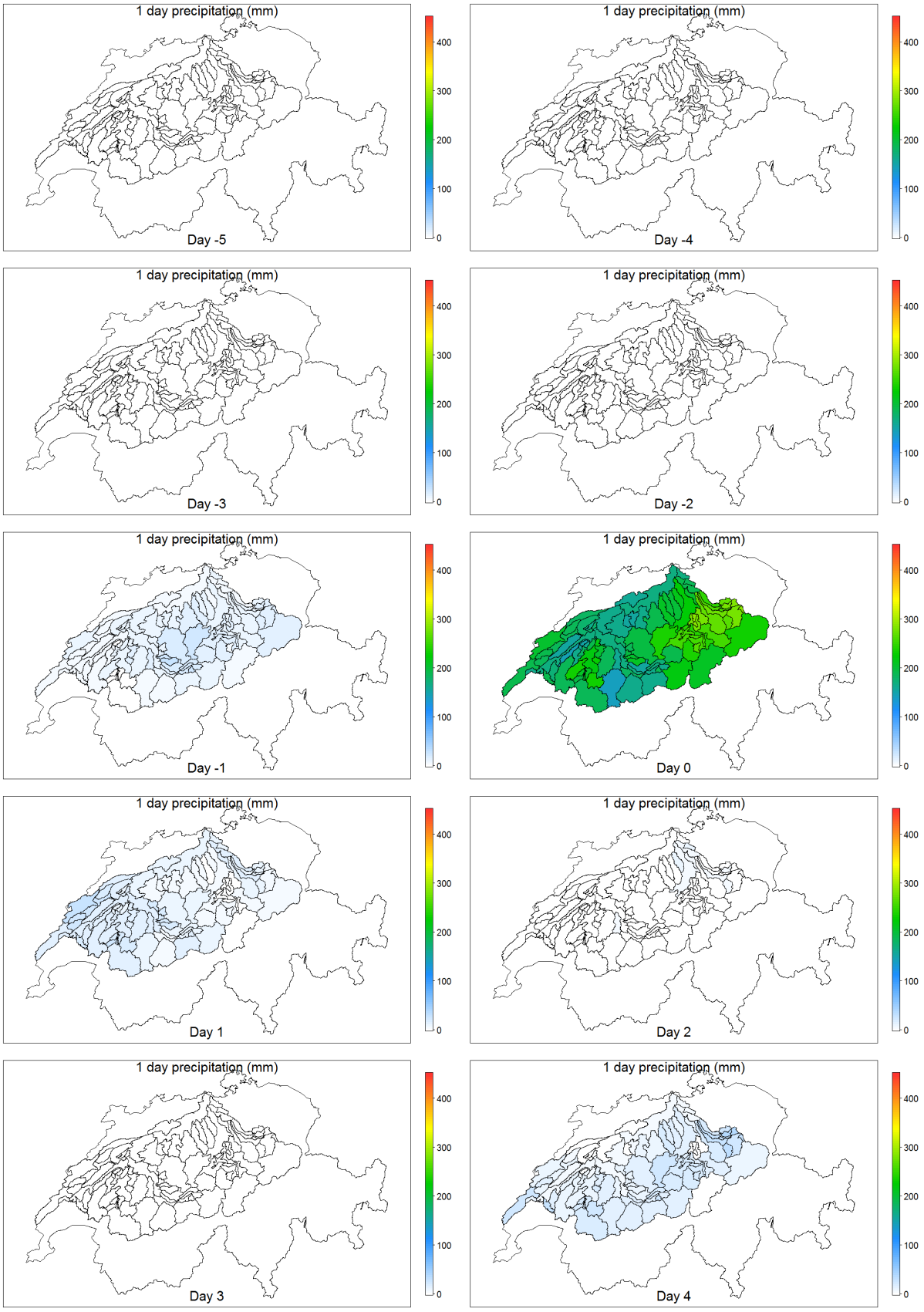


Figure A. 55 Space/time dynamics of precipitation during 10 days for the 1st 1-day-precipitation event.

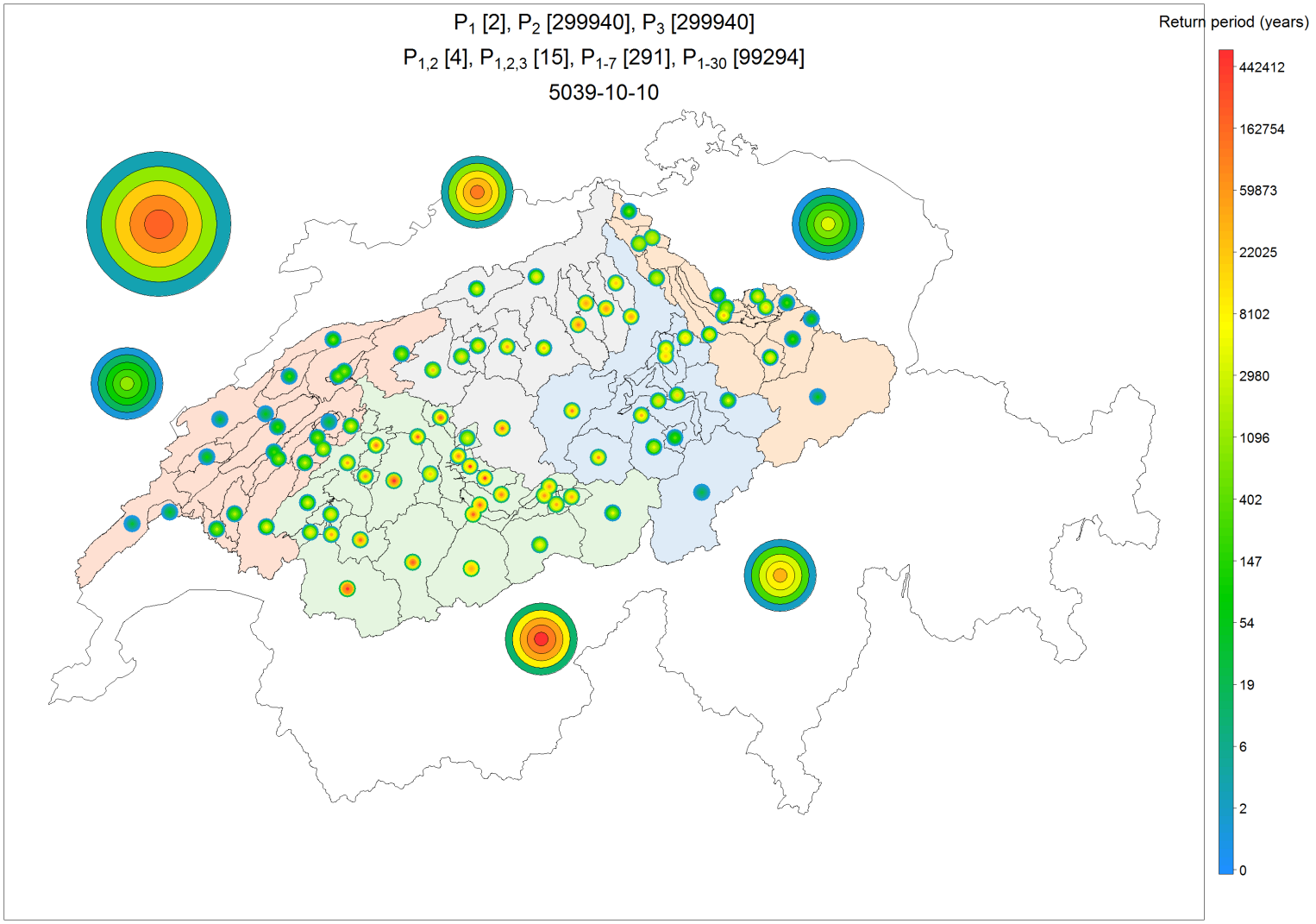


Figure A. 56 Severity map representing the severity at several spatial scales and for several accumulation periods for the 2nd 1-day-precipitation event.

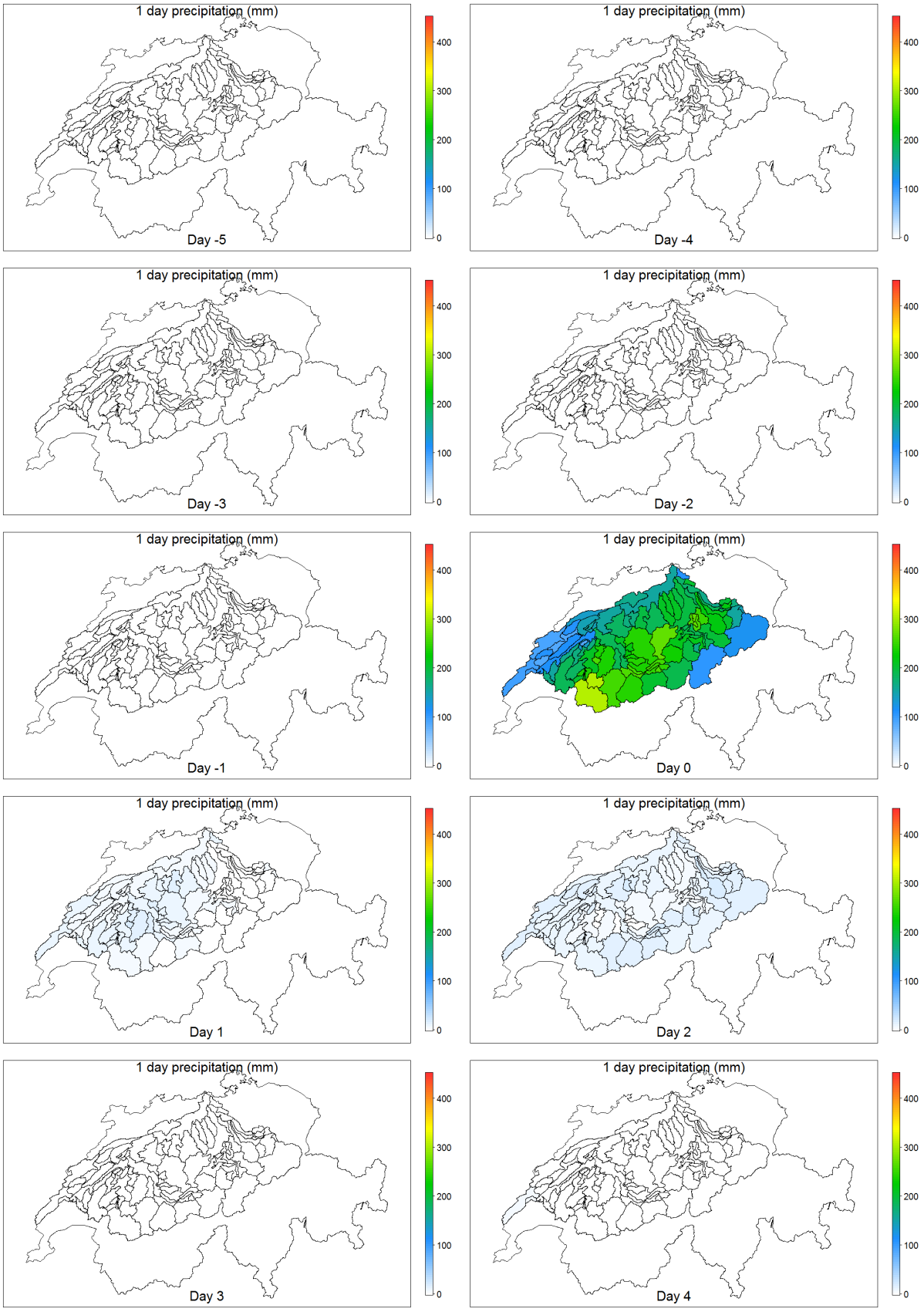


Figure A. 57 Space/time dynamics of precipitation during 10 days for the 2nd 1-day-precipitation event.

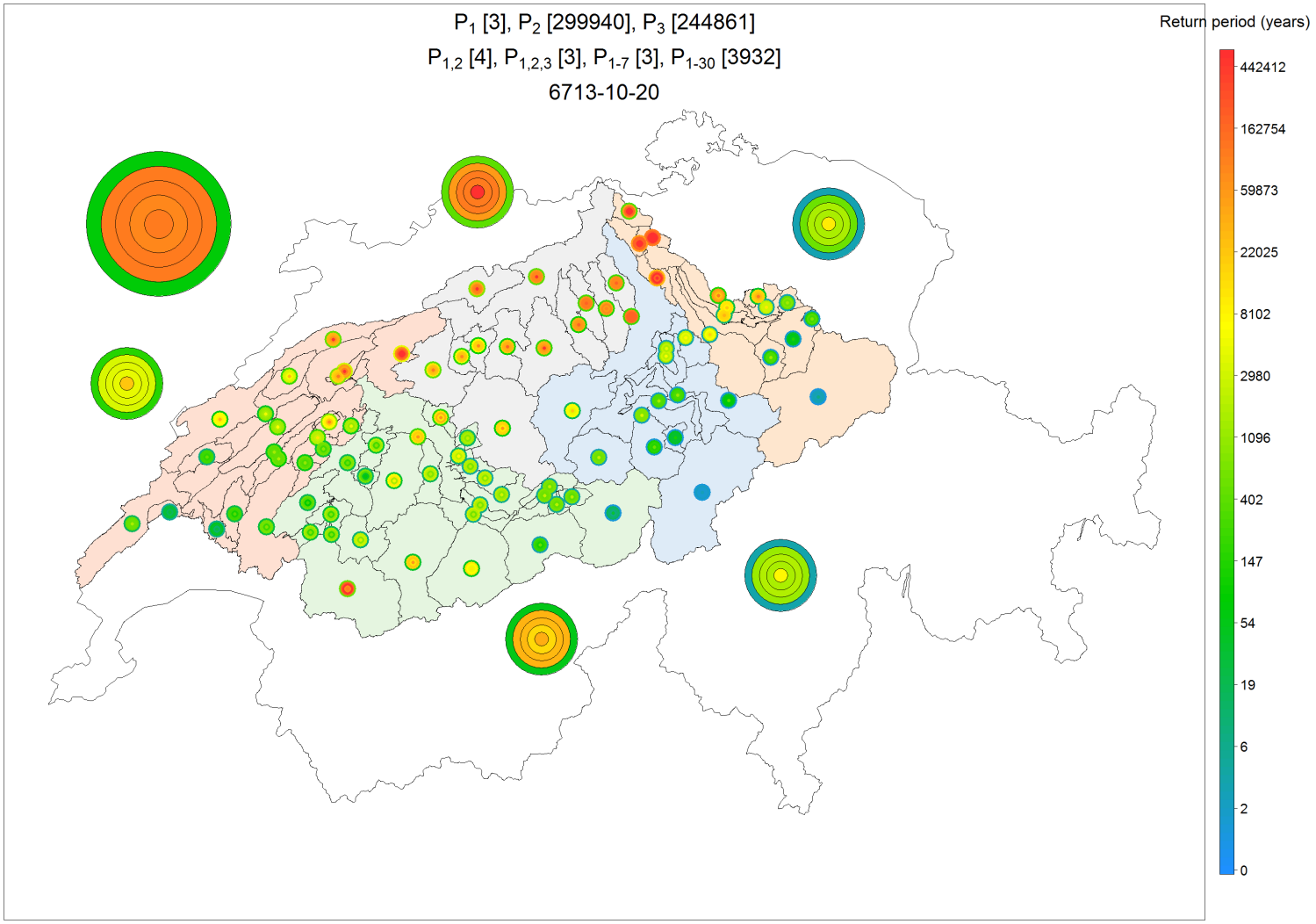


Figure A. 58 Severity map representing the severity at several spatial scales and for several accumulation periods for the 3rd 1-day-precipitation event.

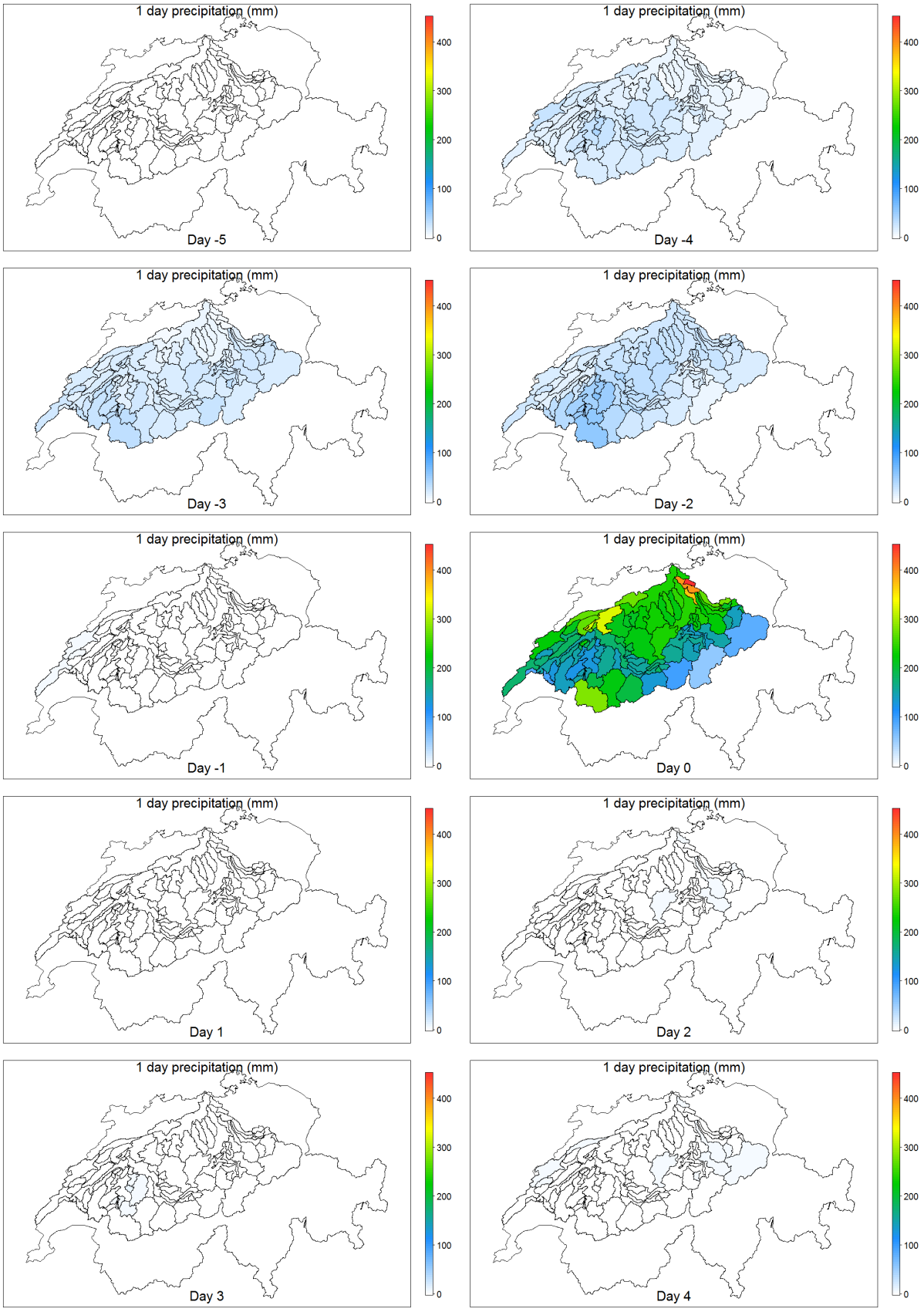


Figure A. 59 Space/time dynamics of precipitation during 10 days for the 3rd 1-day-precipitation event.

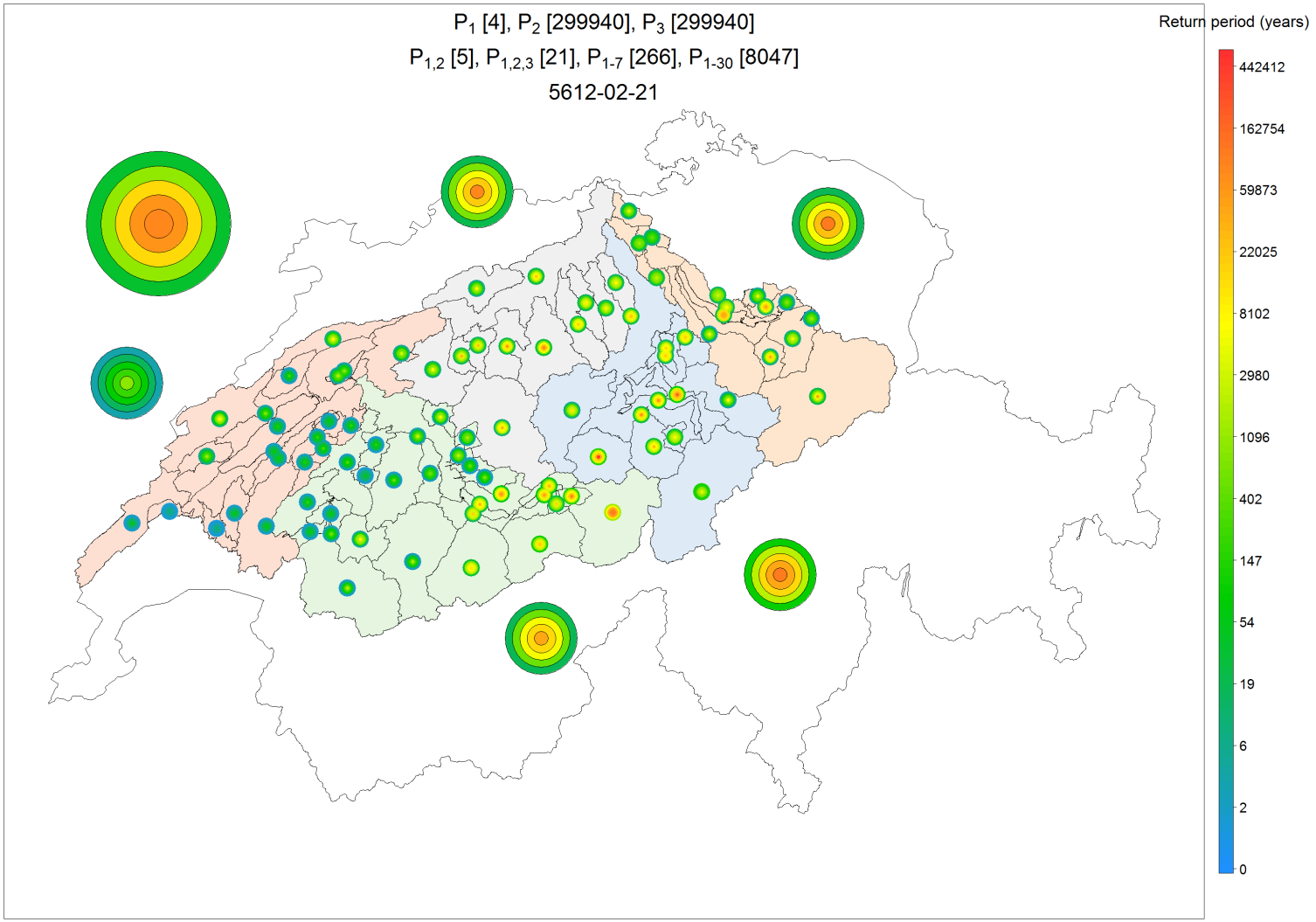


Figure A. 60 Severity map representing the severity at several spatial scales and for several accumulation periods for the 4th1-day-precipitation event

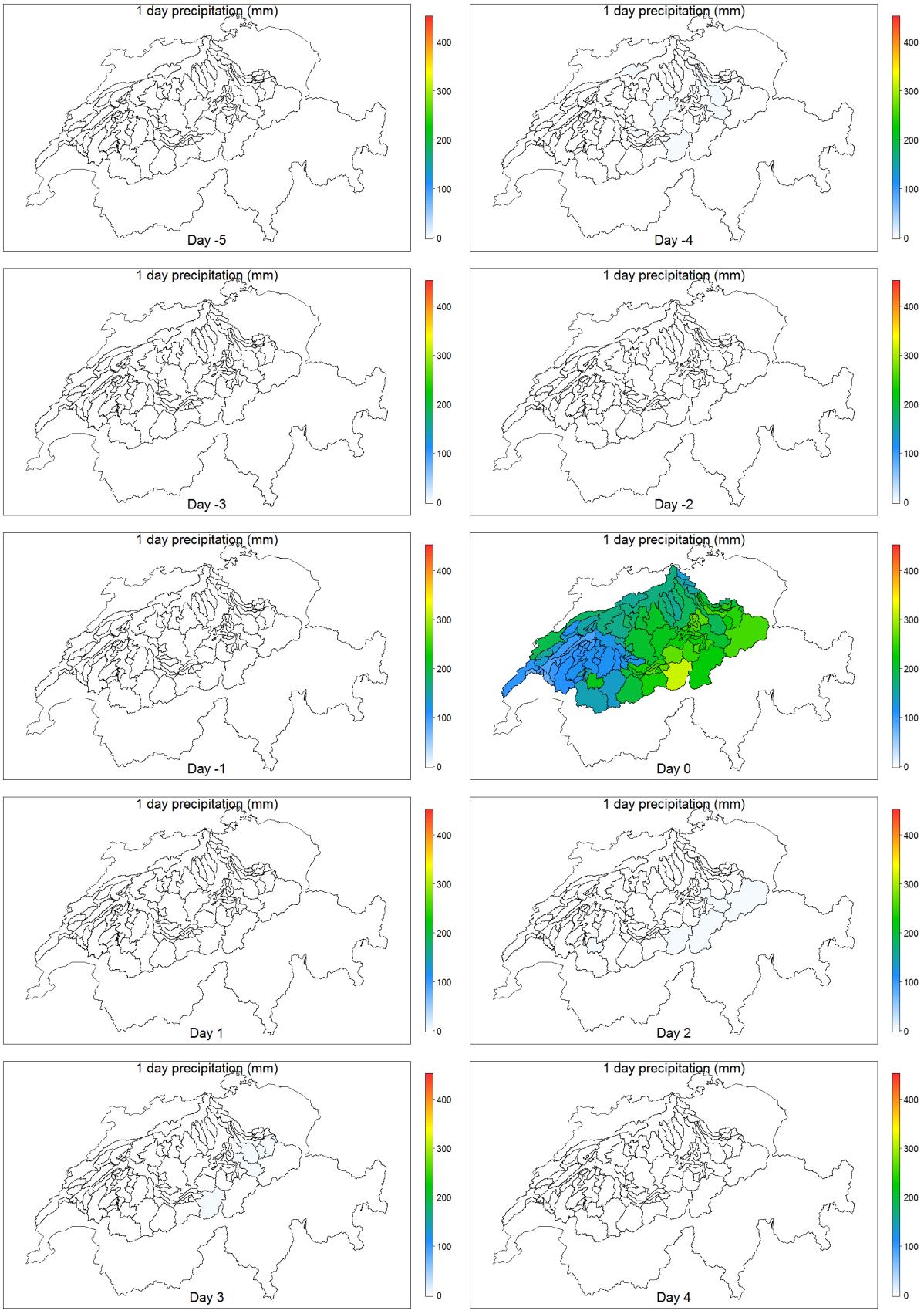


Figure A. 61 Space/time dynamics of precipitation during 10 days for the 4th 1-day-precipitation event.

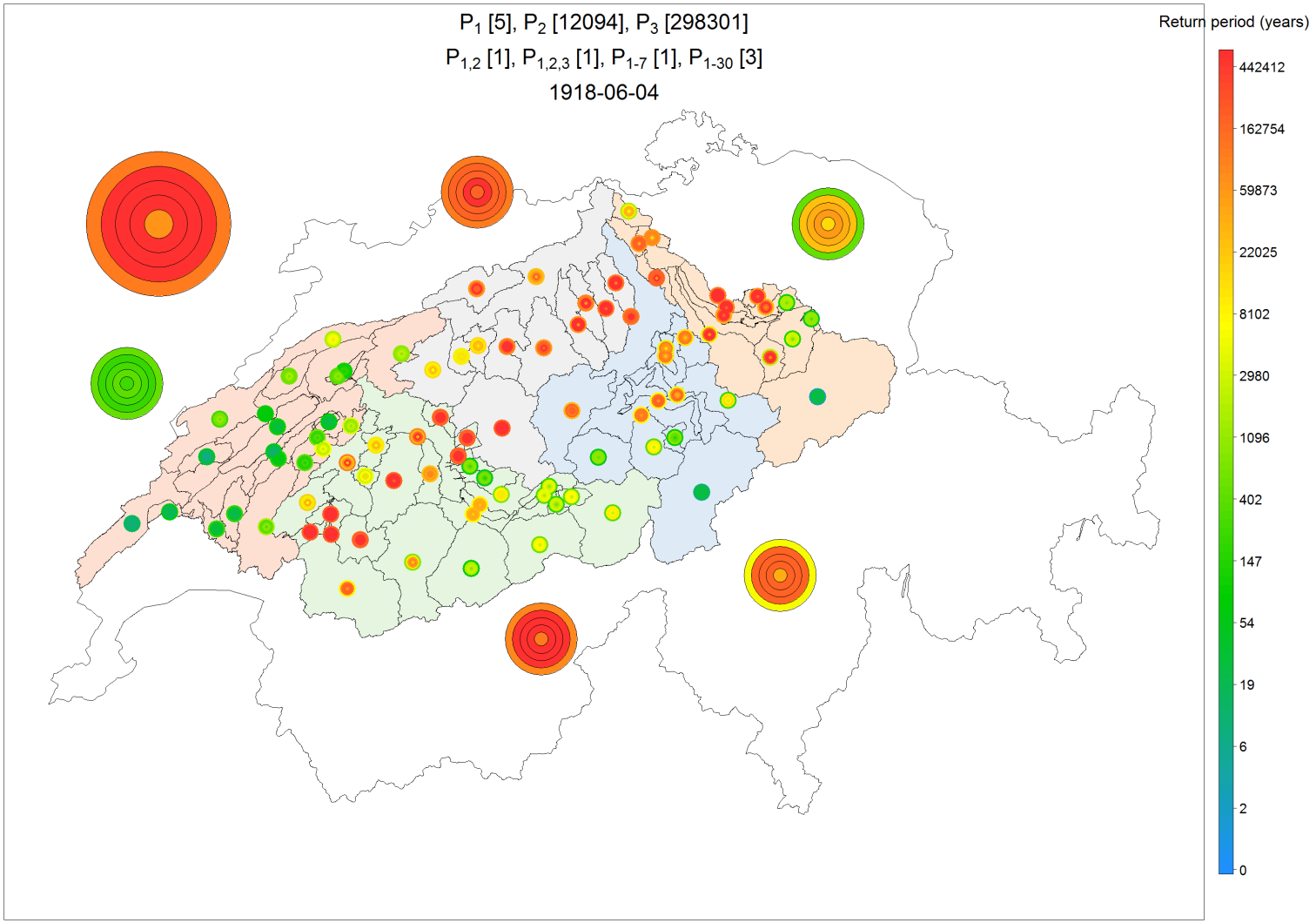


Figure A. 62 Severity map representing the severity at several spatial scales and for several accumulation periods for the 5th1-day-precipitation event.

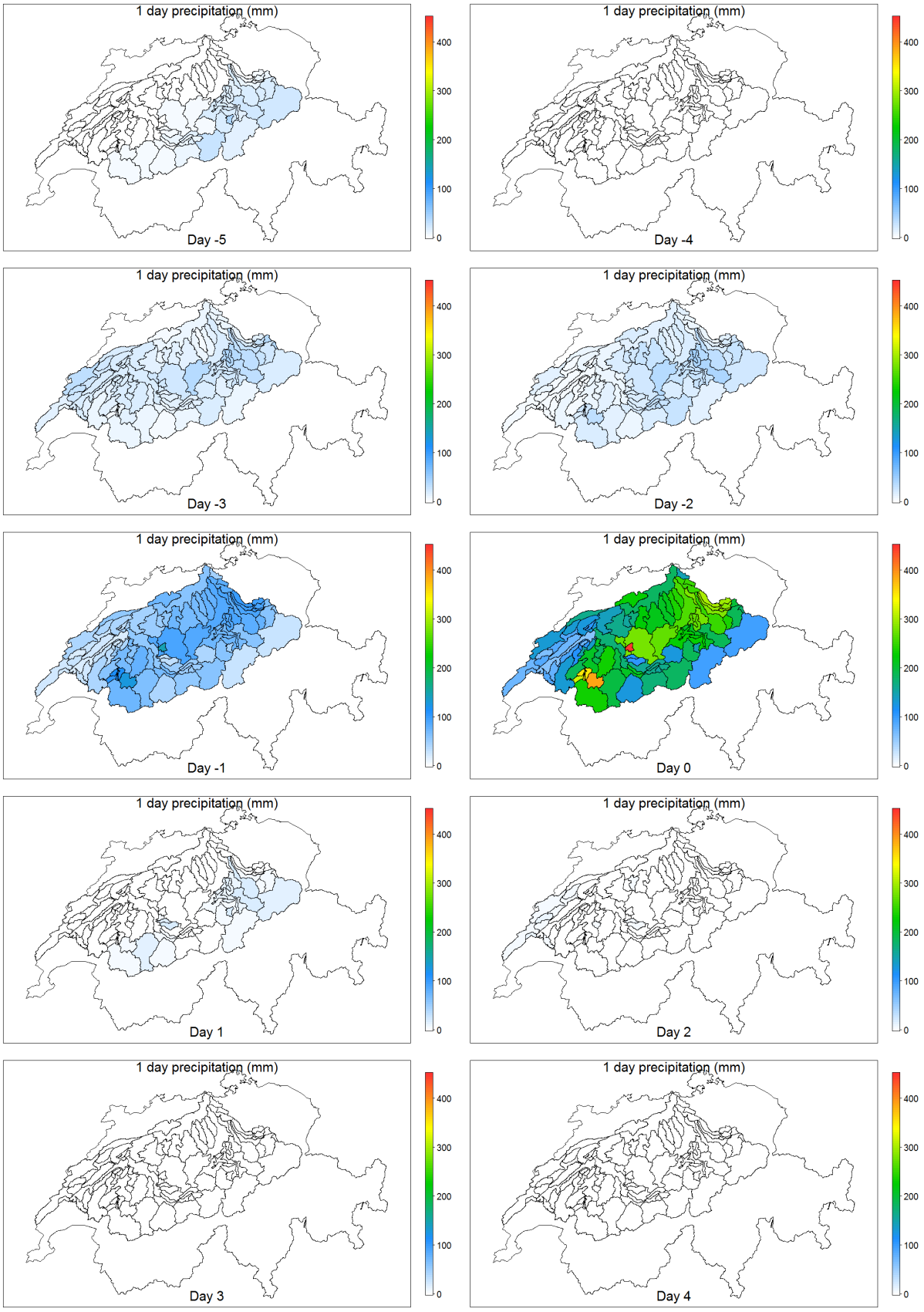


Figure A. 63 Space/time dynamics of precipitation during 10 days for the 5th 1-day-precipitation event.

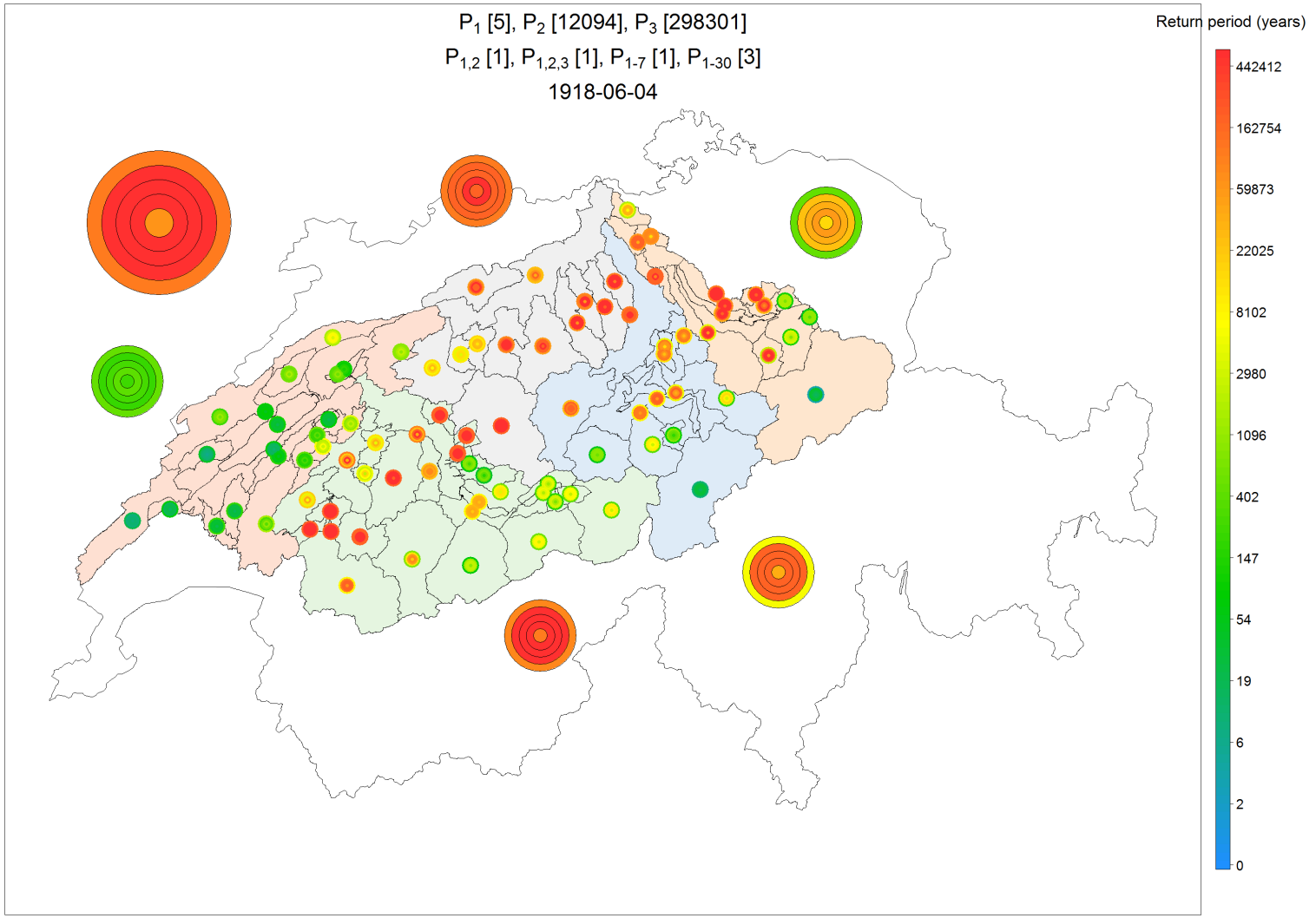


Figure A. 64 Severity map representing the severity at several spatial scales and for several accumulation periods for the 1st 3-day-precipitation event.

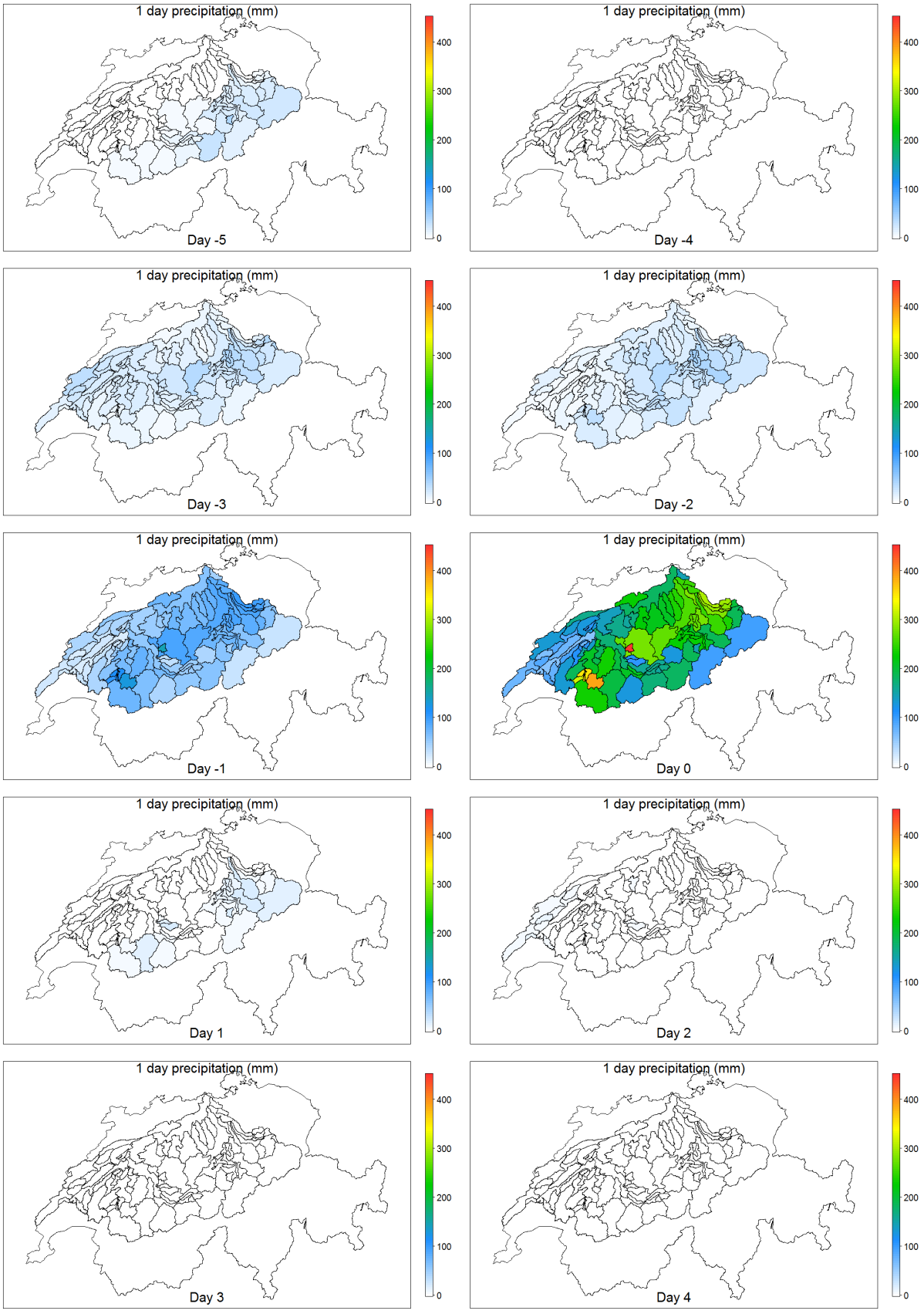


Figure A. 65 Space/time dynamics of precipitation during 10 days for the 1st 3-day-precipitation event.

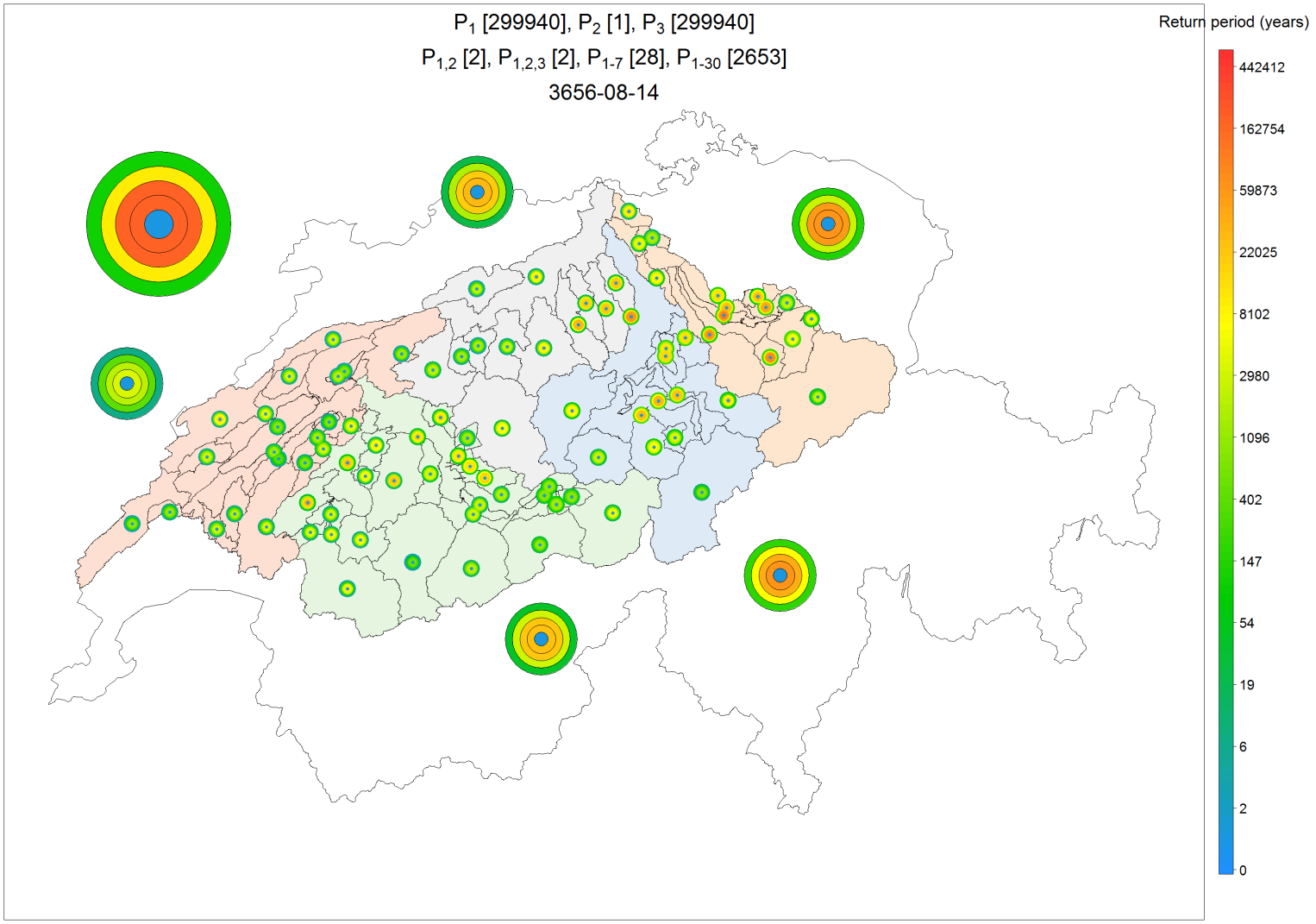


Figure A. 66 Severity map representing the severity at several spatial scales and for several accumulation periods for the 2nd 3-day-precipitation event.

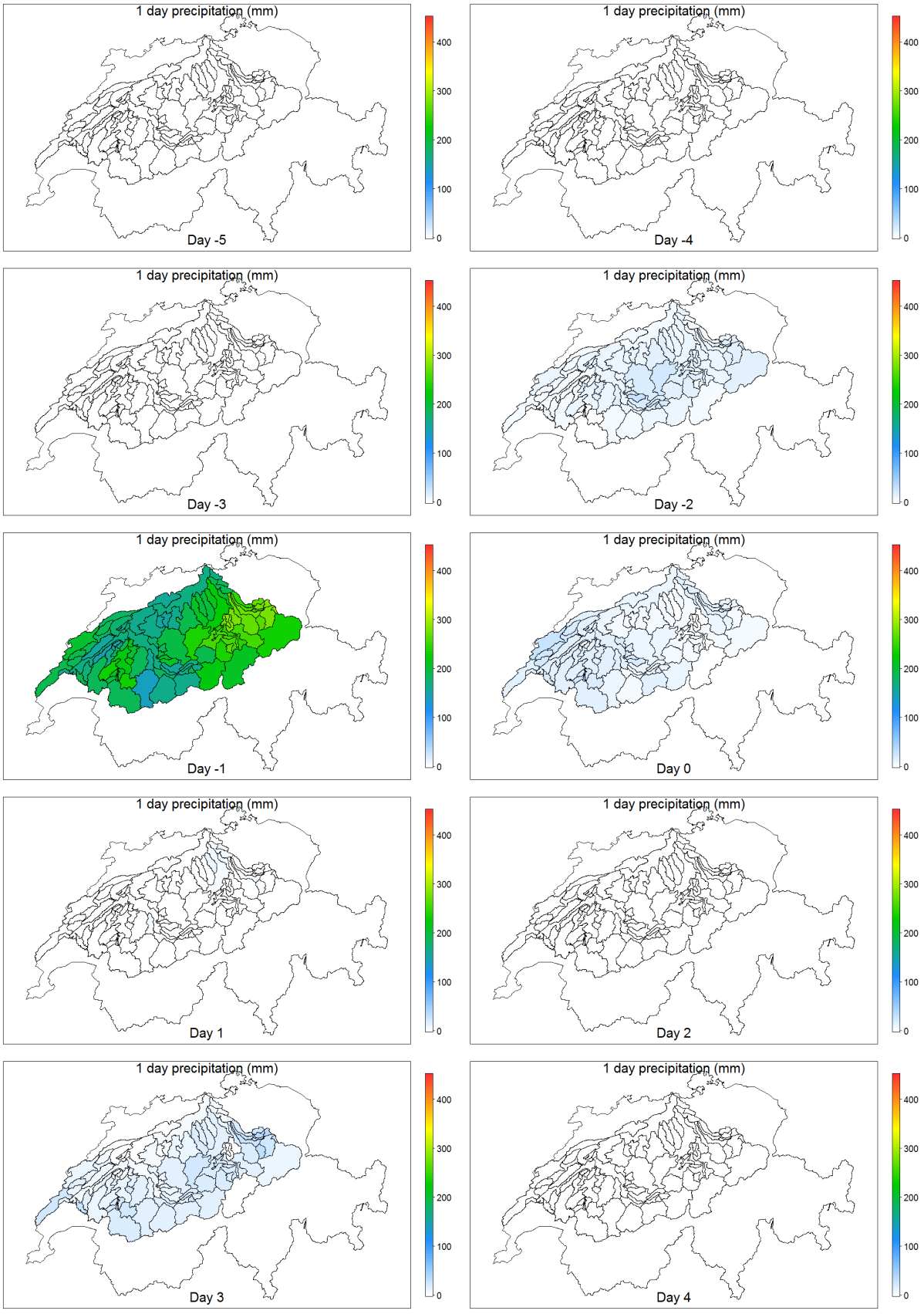


Figure A. 67 Space/time dynamics of precipitation during 10 days for the 2nd 3-day-precipitation event.

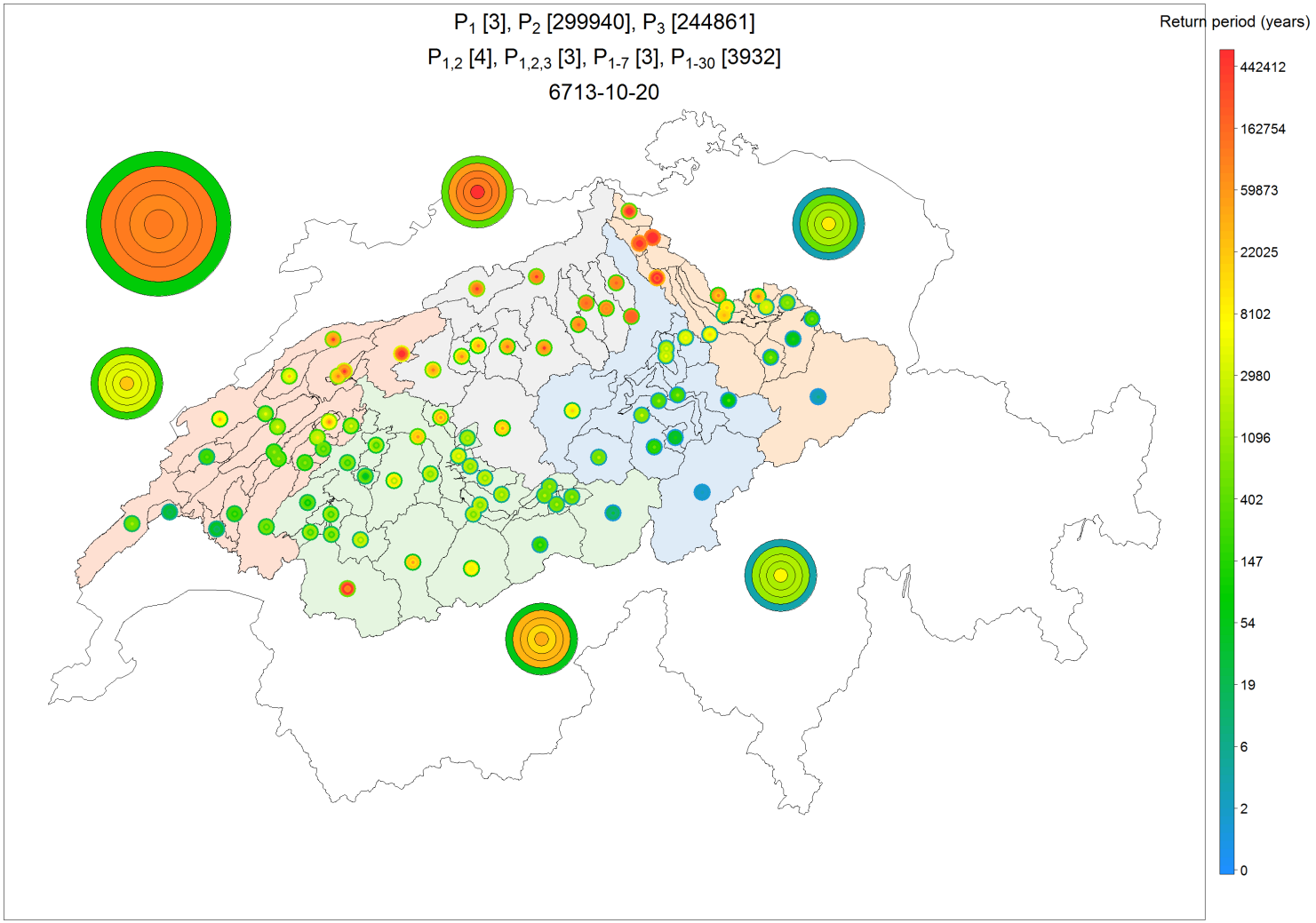


Figure A. 68 Severity map representing the severity at several spatial scales and for several accumulation periods for the 3rd 3-day-precipitation event.

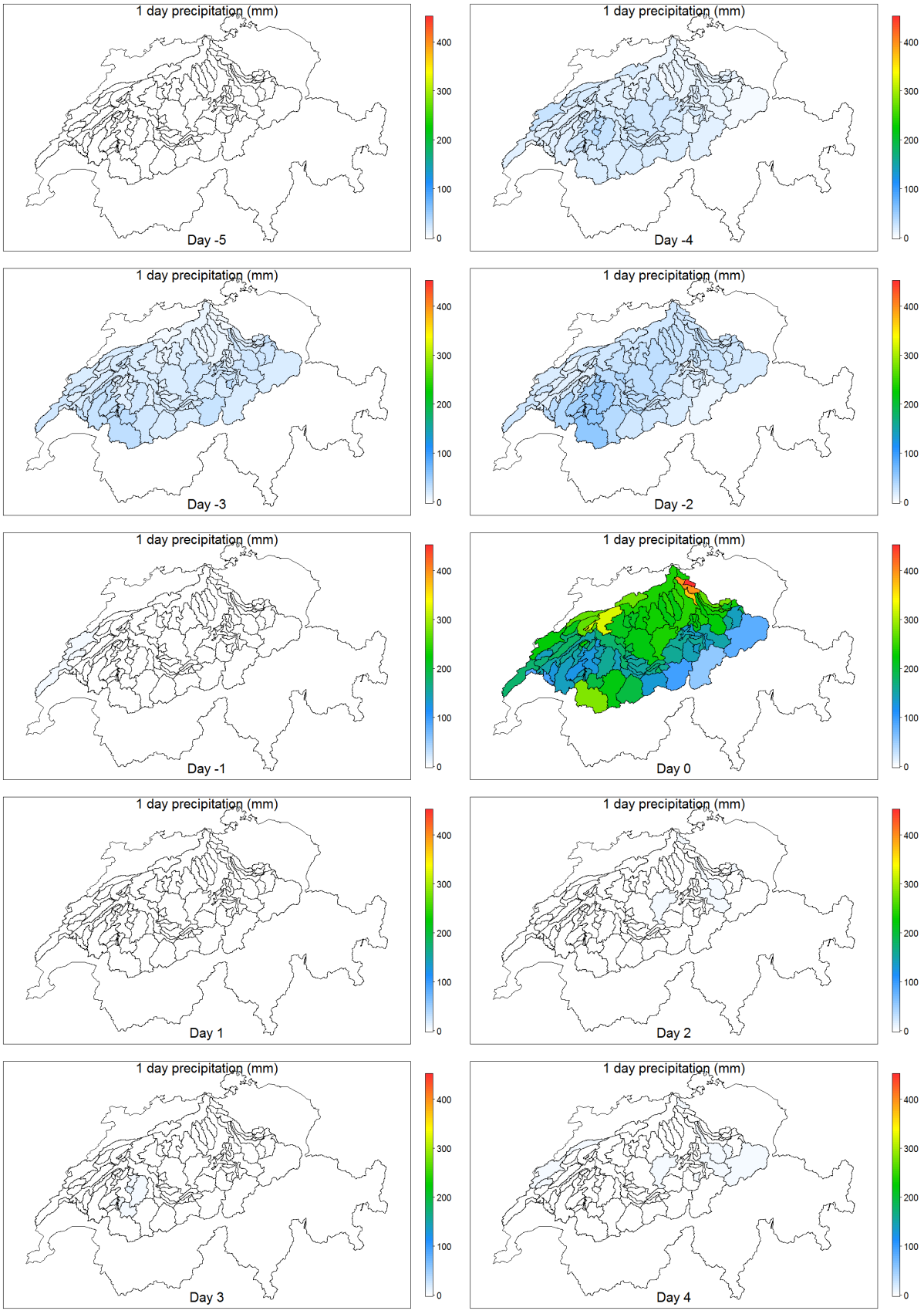


Figure A. 69 Space/time dynamics of precipitation during 10 days for the3rd 3-day-precipitation event.

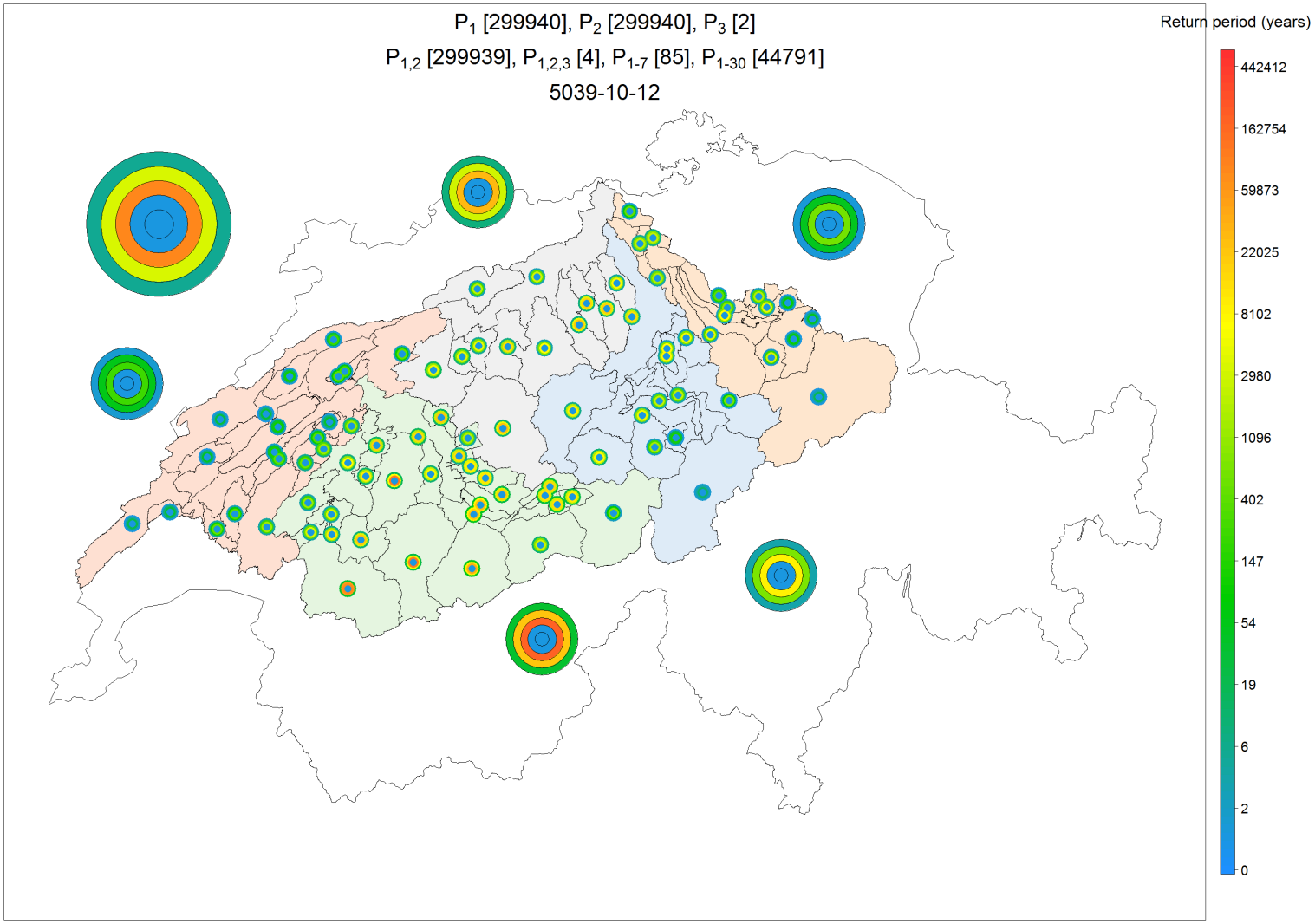


Figure A. 70 Severity map representing the severity at several spatial scales and for several accumulation periods for the 4th 3-day-precipitation event.

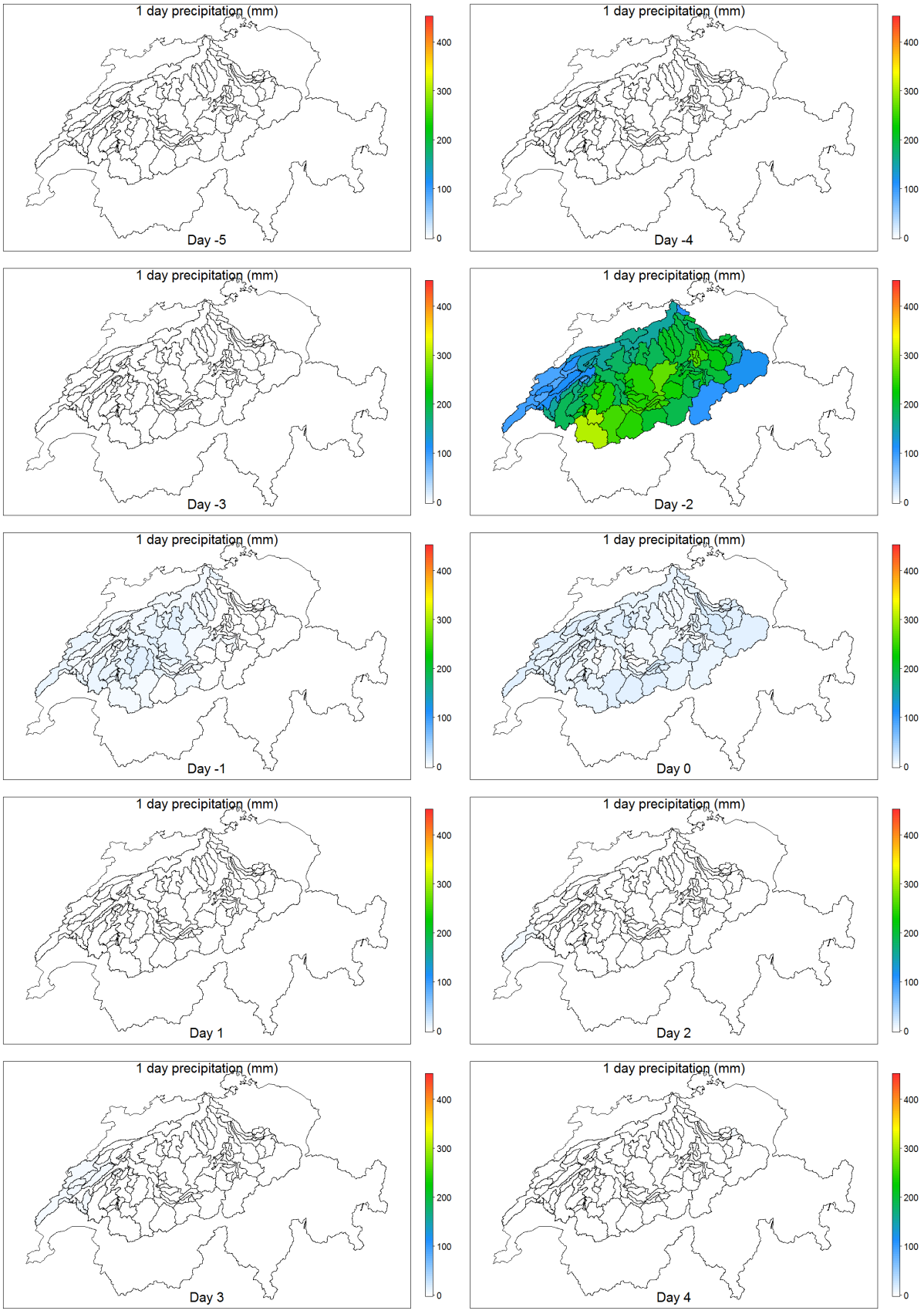


Figure A. 71 Space/time dynamics of precipitation during 10 days for the 4th 3-day-precipitation event.

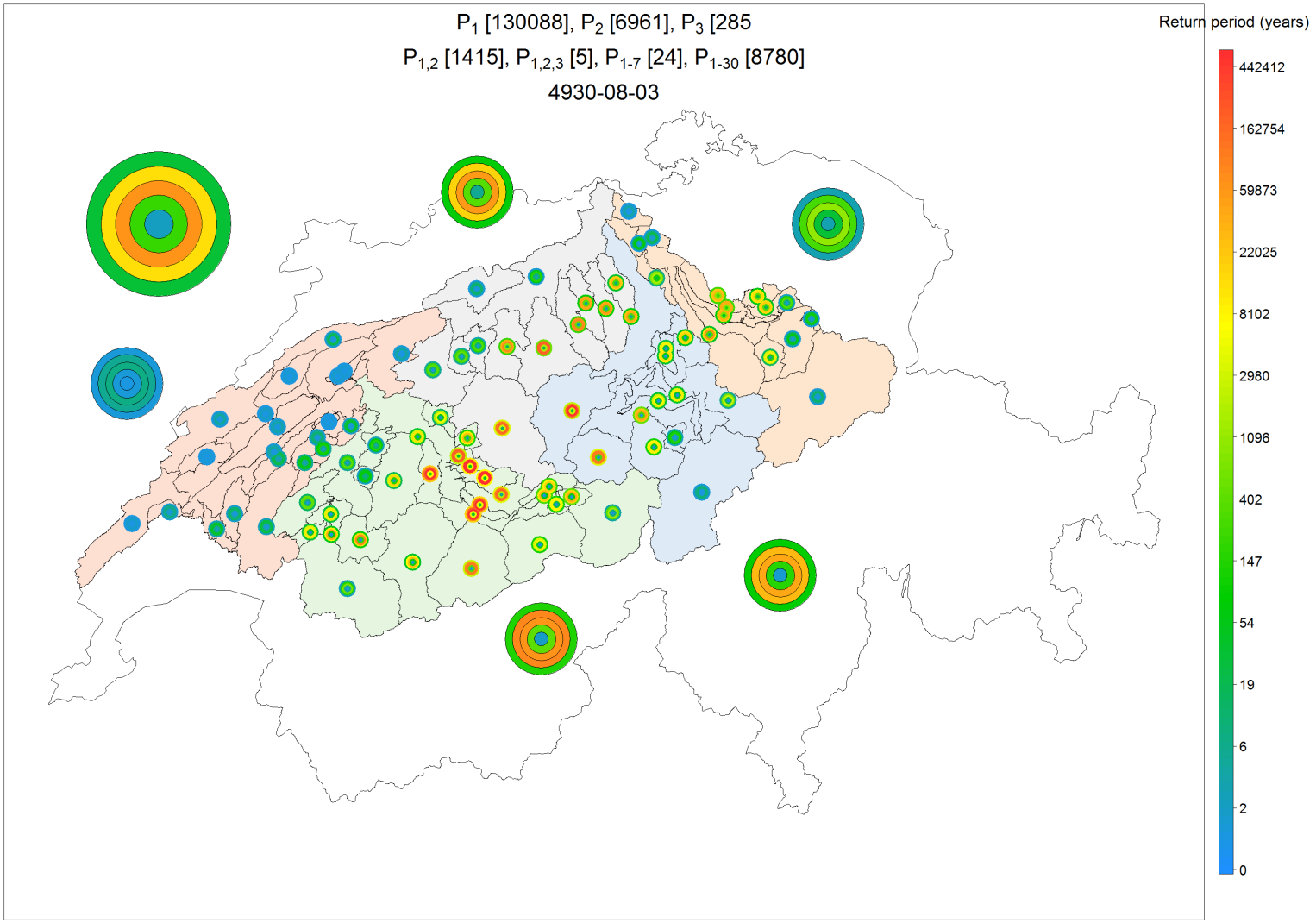


Figure A. 72 Severity map representing the severity at several spatial scales and for several accumulation periods for the 5th 3-day-precipitation event.

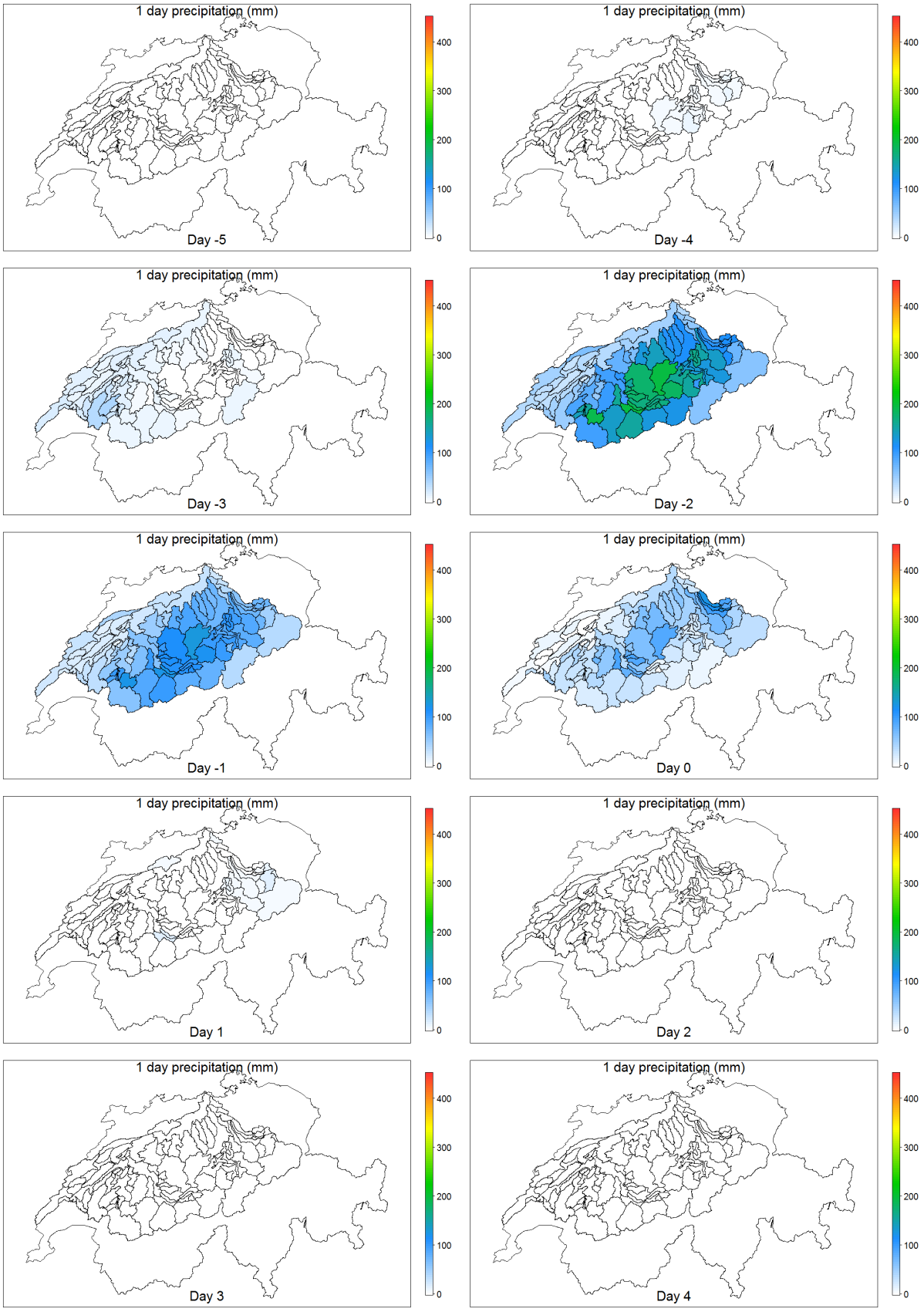


Figure A. 73 Space/time dynamics of precipitation during 10 days for the 5th 3-day-precipitation event.

Largest hydrological events with SCAMP: Severity maps and space/time dynamics The precipitation space-time patterns during the largest hydrological events are shown in the following figures: for each event first the severity map is shown, then the space/time dynamics of precipitation. The severity map represents the severity of this event at several spatial scales and for several accumulation periods; the space/time dynamics of precipitation for 10 days are composed of 10 maps ("Day 0" is the day of the discharge peak).

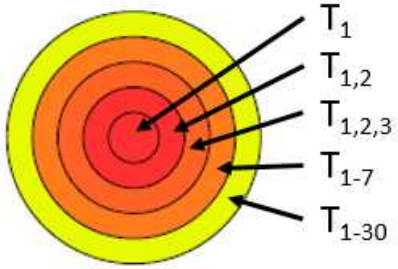
The title of the severity map shows the rank (in descending order) of the precipitation events at the whole catchment scale for several accumulation durations:

- H[]: rank of the hydrological event
- P_0[]: rank of the 1-day precipitation during the day of the hydrological peak (“Day 0”)
- P_1[]: rank of the 1-day precipitation during the day before the hydrological peak (“Day -1”)
- P_2[]: rank of the 1-day precipitation 2 days before the hydrological peak (“Day -2”)
- P_3[]: rank of the 1-day precipitation 3 days before the hydrological peak (“Day -3”)
- $P_{1,2}$[]: rank of the 2-day precipitation from “Day -1” to “Day -2”
- $P_{1,2,3}$[]: rank of the 3-day precipitation from “Day -1” to “Day -3”
- $P_{1\text{-}7}$[]: rank of the 7-day precipitation from “Day -1” to “Day -7”
- $P_{1\text{-}30}$[]: rank of the 30-day precipitation from “Day -1” to “Day -30”
- In addition, the fictive date of the hydrological peak is indicated on the third line

The severity of the simulated event is then represented with a set of severity-board. Each severity-board gives the return period of precipitation simulated for this event when cumulated over several durations (all ending the day before the hydrological peak):

- 1 day (center of the severity-board)
- 2 days
- 3 days
- 7 days
- 30 days (edge of the severity-board)

For instance, if T_1, $T_{1,2}$, $T_{1,2,3}$, $T_{1\text{-}7}$, and $T_{1\text{-}30}$ are the return periods associated with the precipitation P_1, $P_{1,2}$, $P_{1,2,3}$, $P_{1\text{-}7}$, and $P_{1\text{-}30}$, respectively



The color scale for the severity-boards is given in the right-hand side of the figure. Severity-boards are given for precipitation simulated over different spatial scales:

- local precipitation at each of the 89 simulation sub-catchments (small circles)
- mean areal precipitation over each of the 5 main sub-catchments (medium circles)
- whole catchment (large circle)

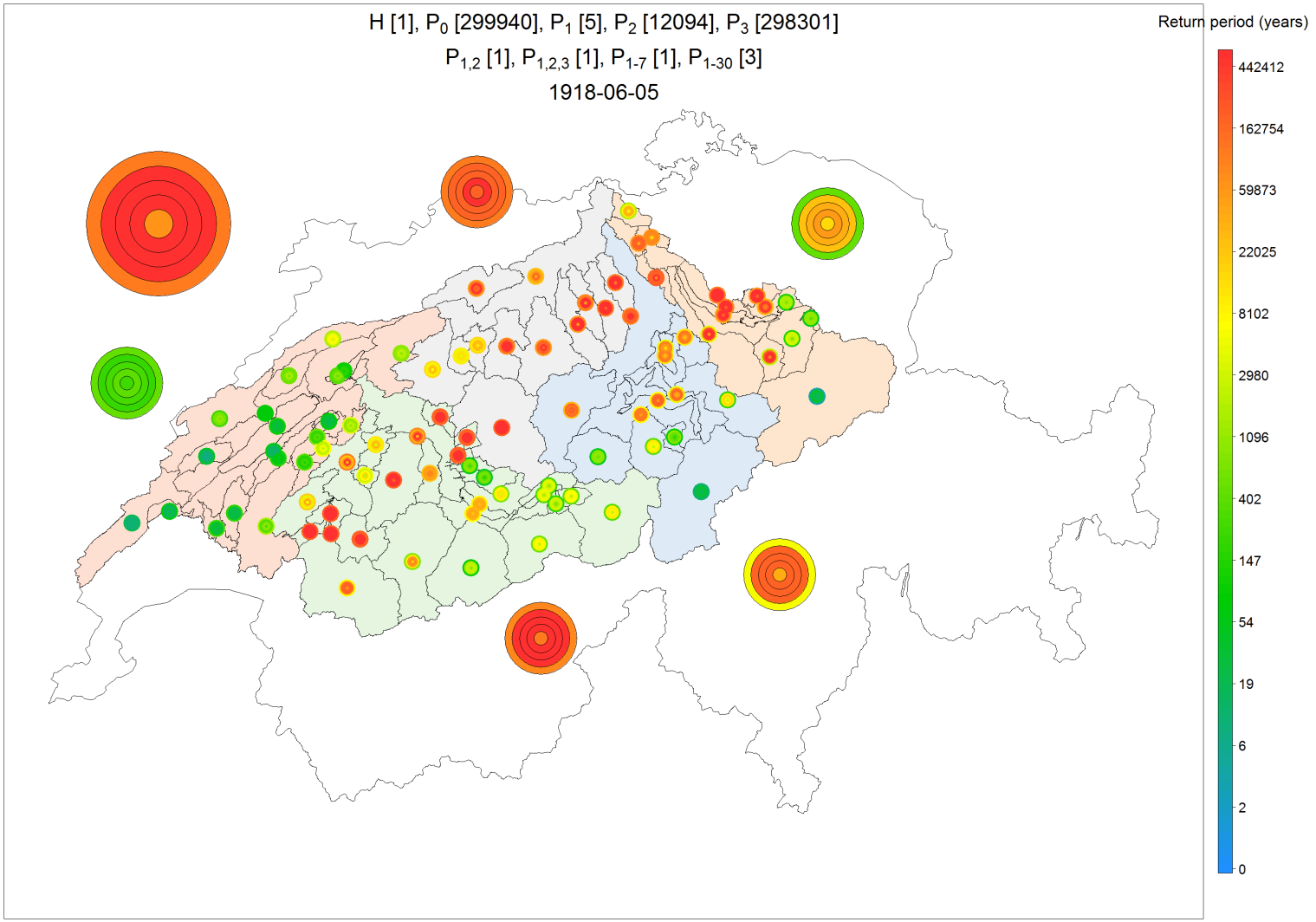


Figure A. 74 Severity map representing the severity at several spatial scales and for several accumulation periods for the 1st hydrological event

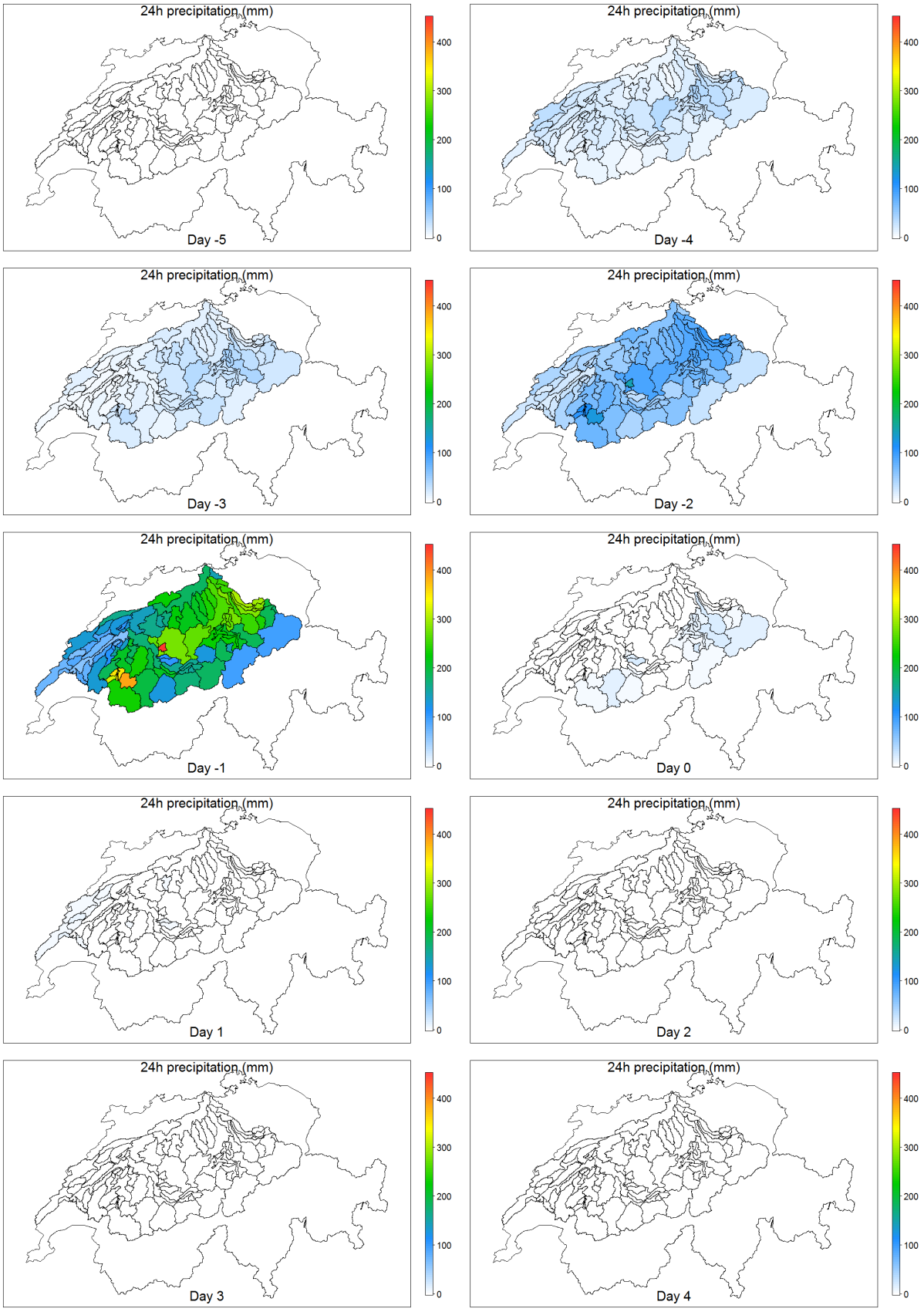


Figure A. 75 Space/time dynamics of precipitation during 10 days for the1st hydrological event

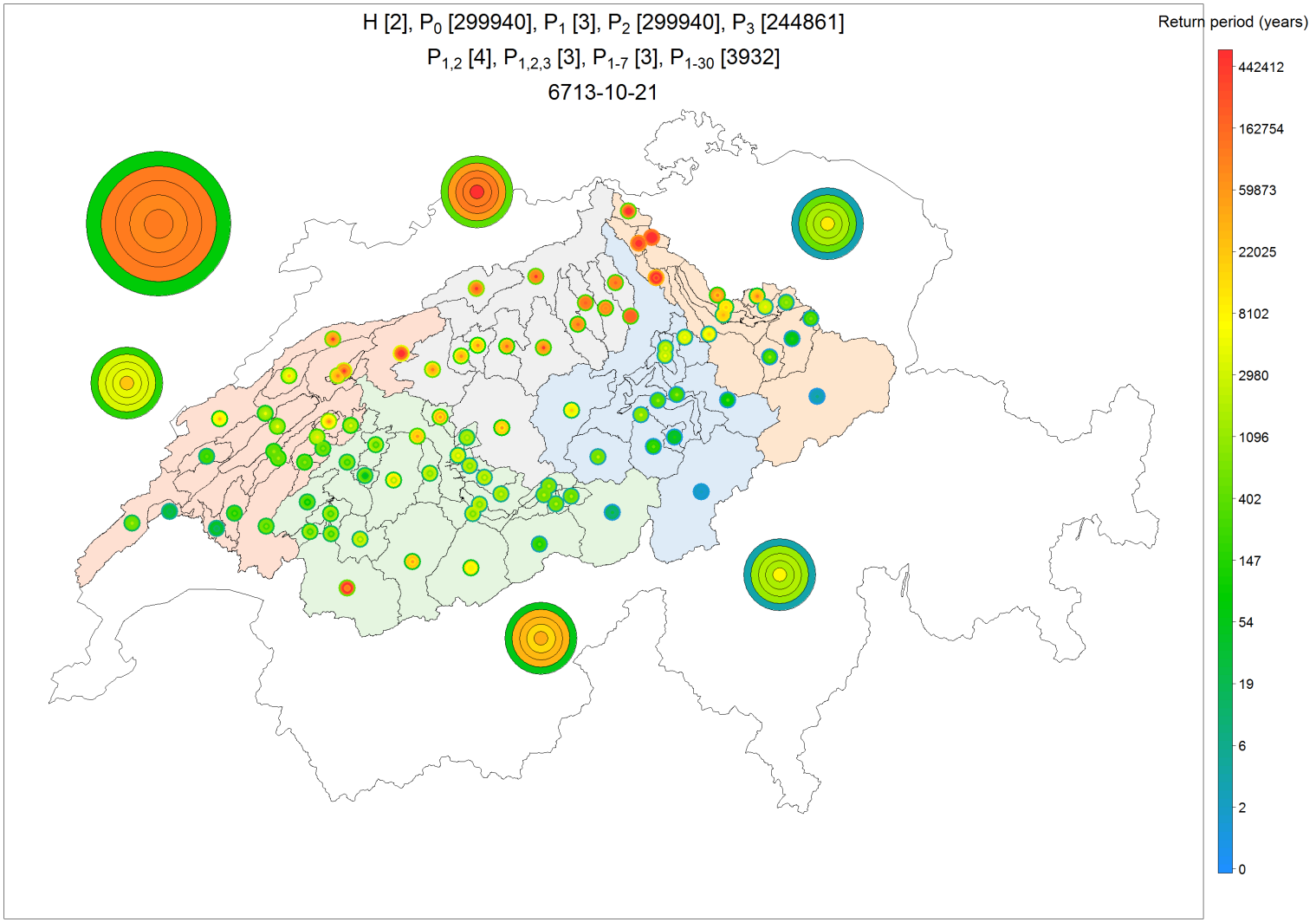


Figure A. 76 Severity map representing the severity at several spatial scales and for several accumulation periods for the 2nd hydrological event

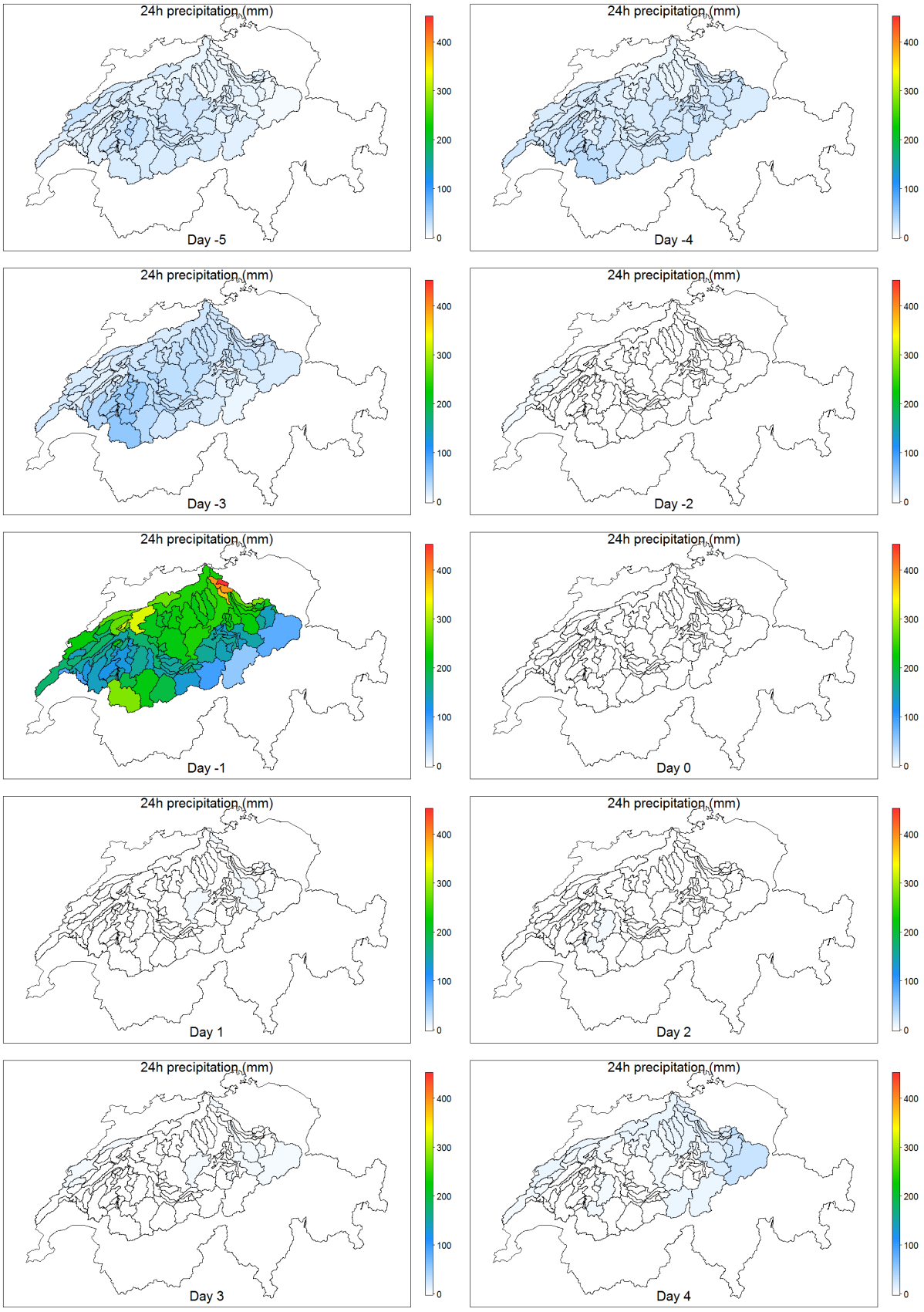


Figure A. 77 Space/time dynamics of precipitation during 10 days for the 2nd hydrological event

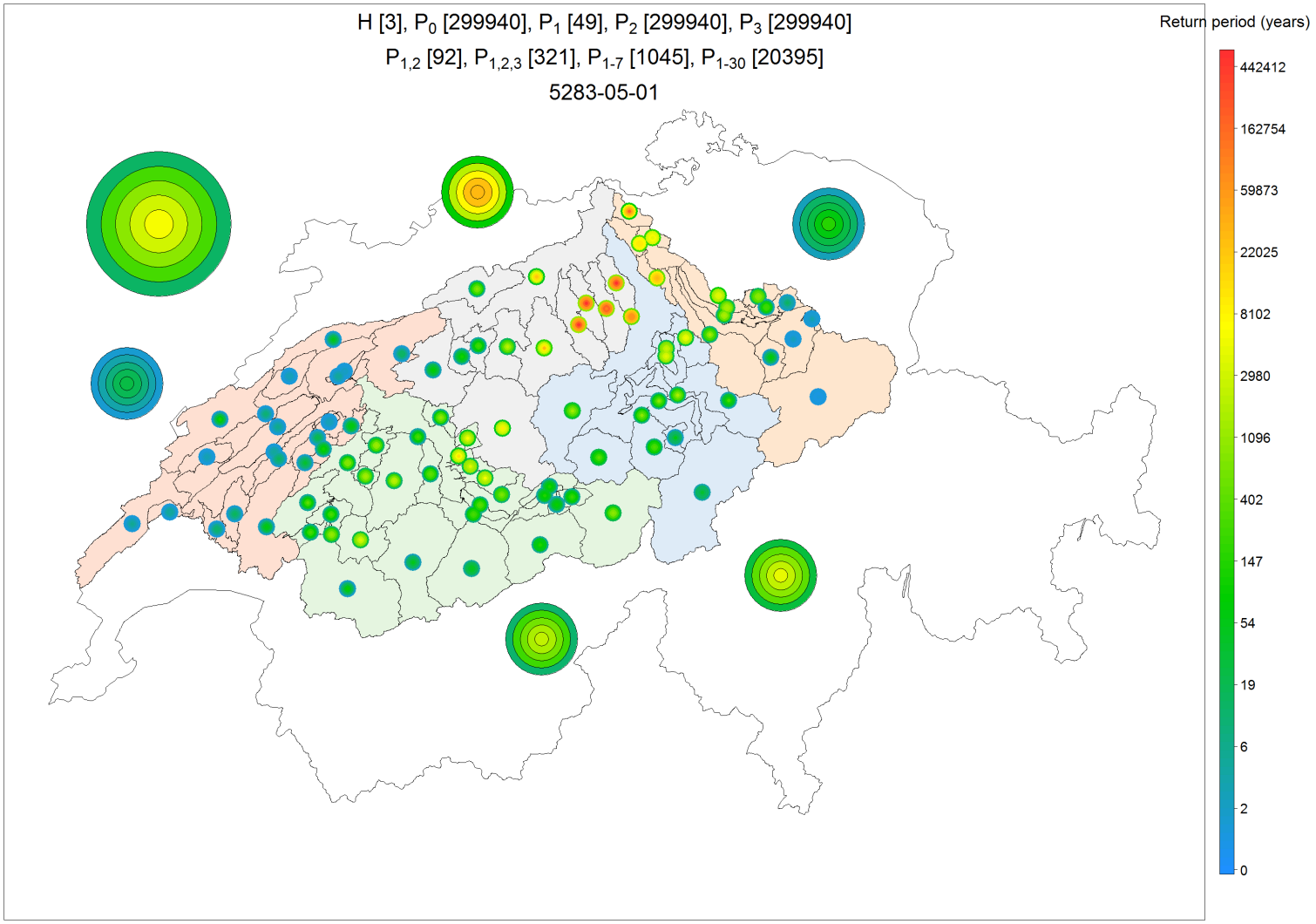


Figure A. 78 Severity map representing the severity at several spatial scales and for several accumulation periods for the 3rd hydrological event

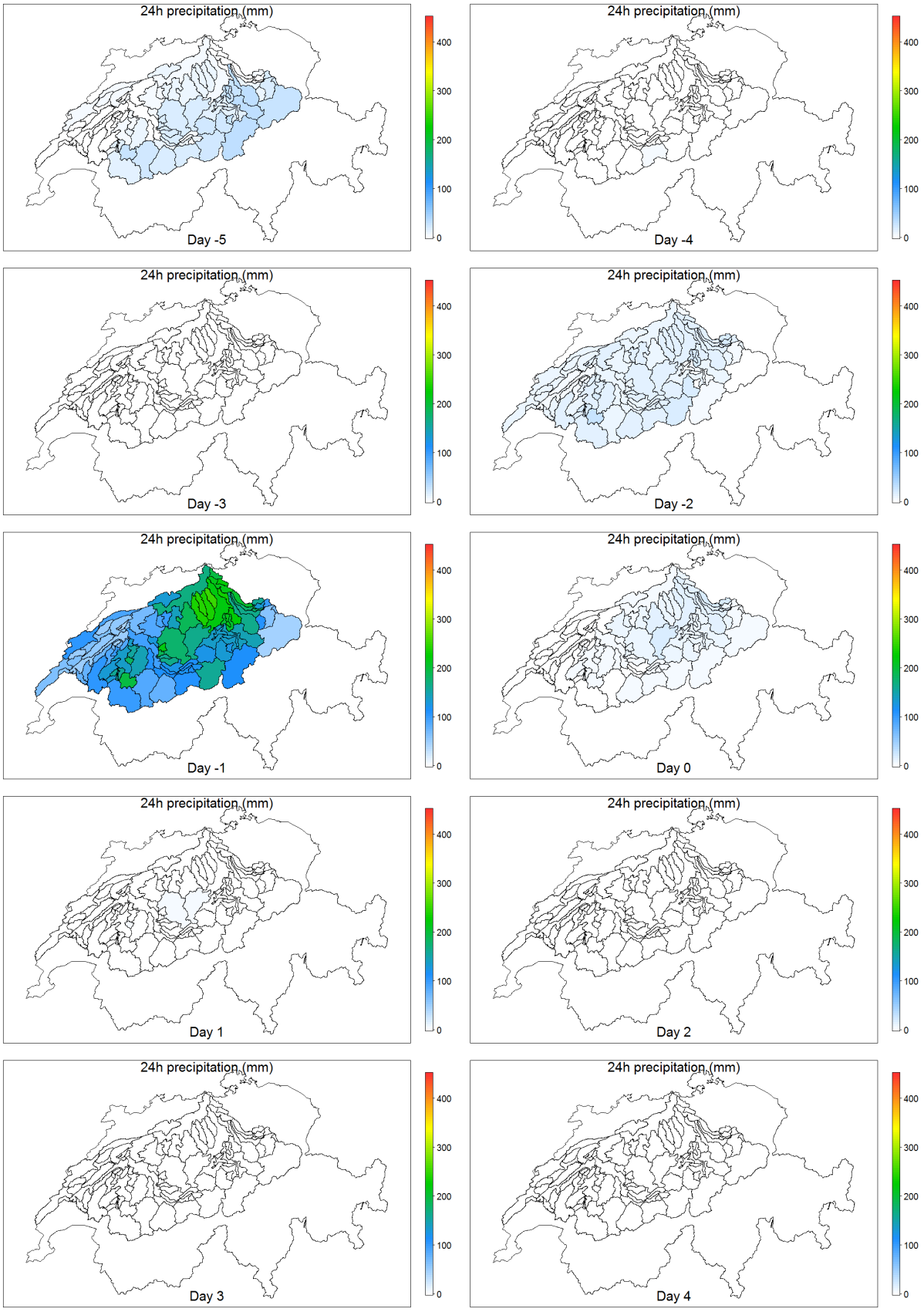


Figure A. 79 Space/time dynamics of precipitation during 10 days for the 3rdhydrological event

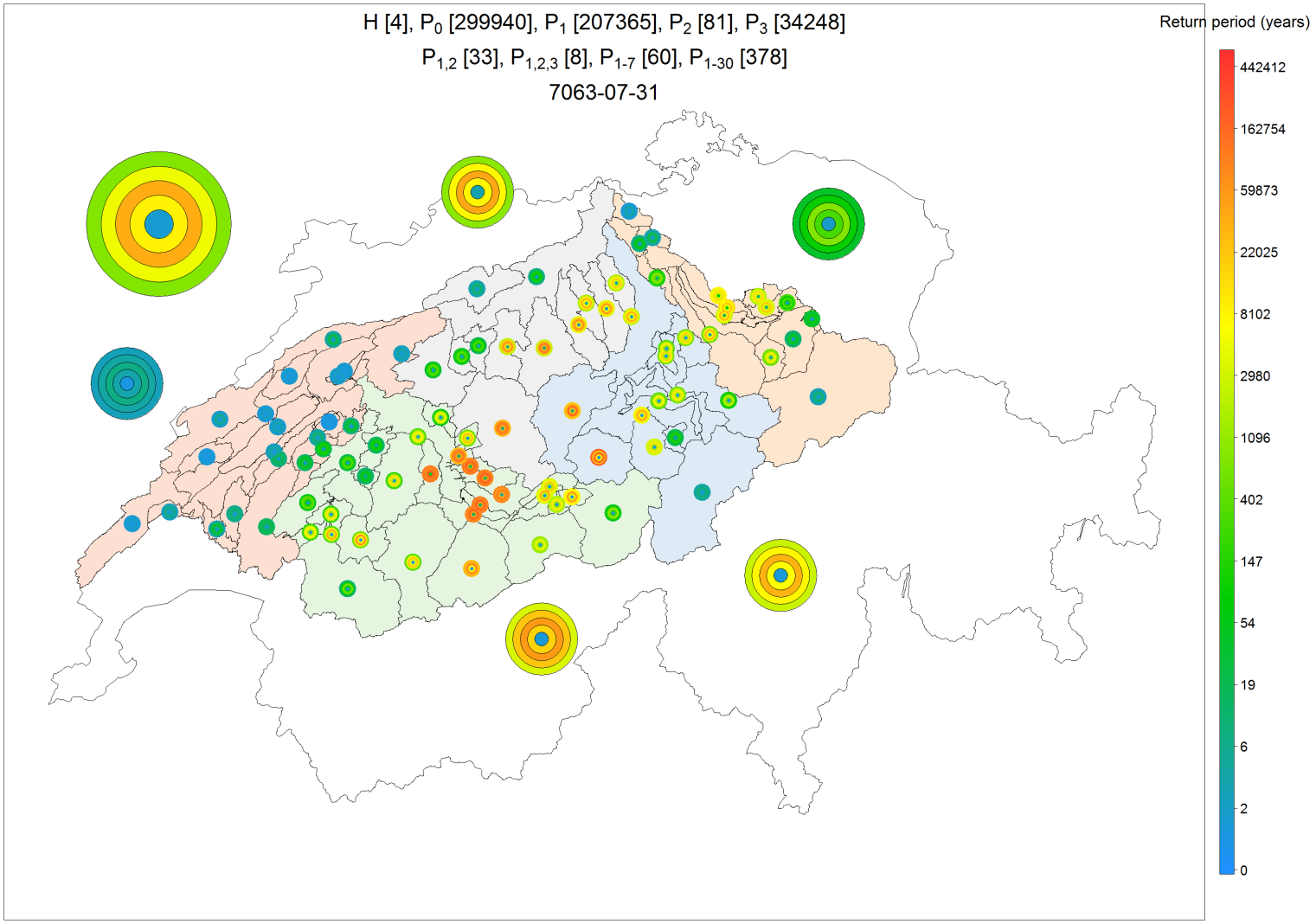


Figure A. 80 Severity map representing the severity at several spatial scales and for several accumulation periods for the 4th hydrological event

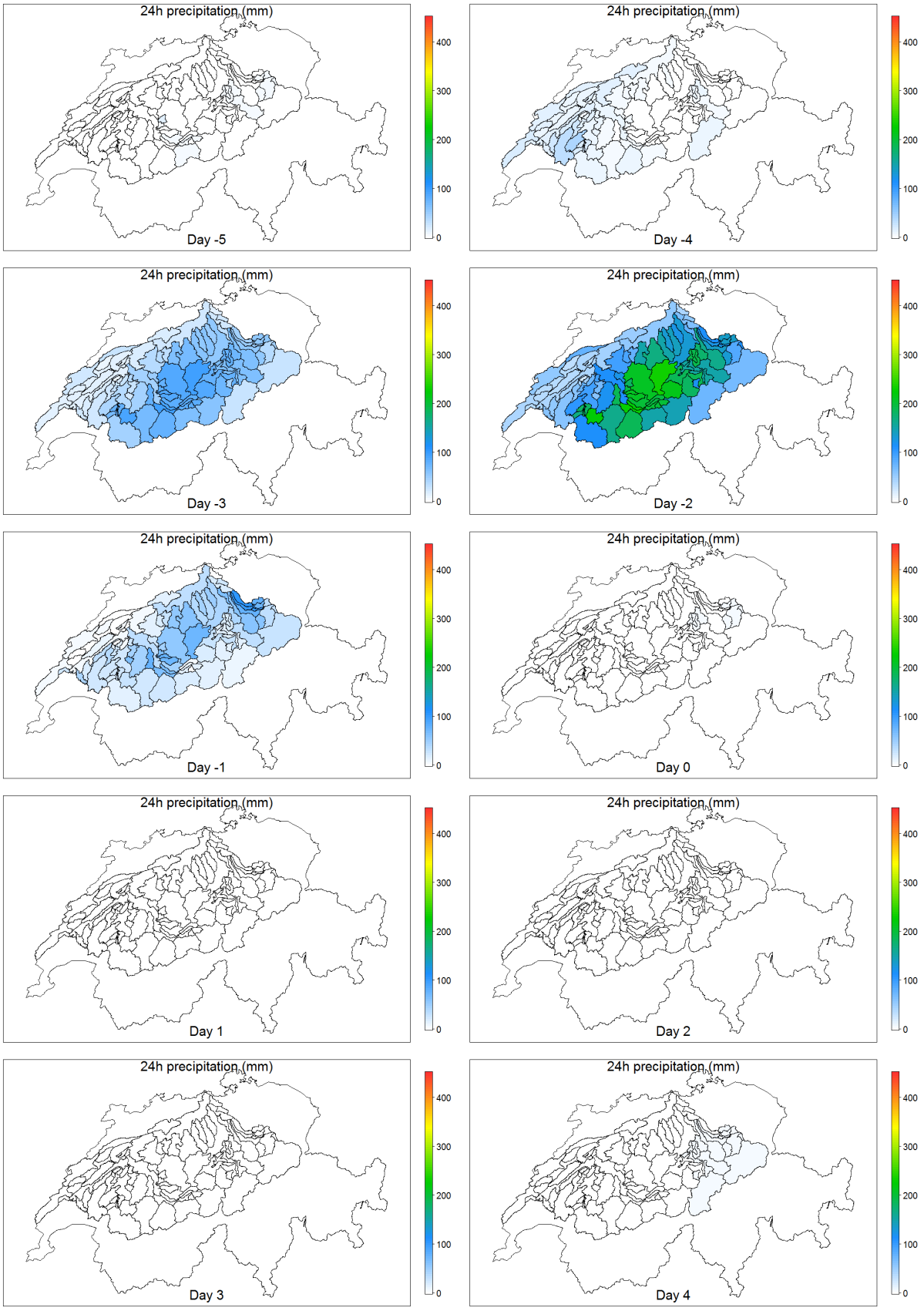


Figure A. 81 Space/time dynamics of precipitation during 10 days for the 4th t hydrological event

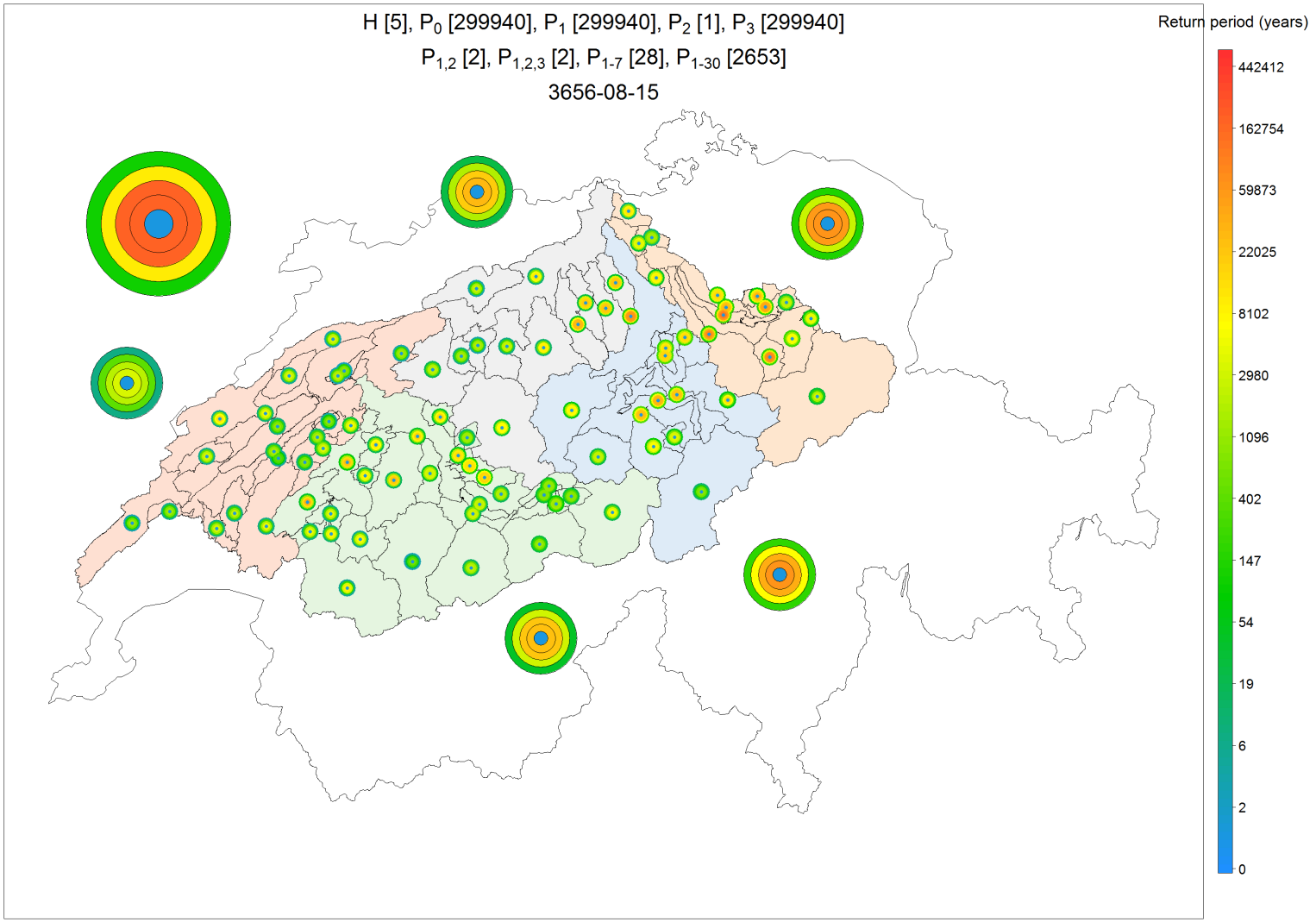


Figure A. 82 Severity map representing the severity at several spatial scales and for several accumulation periods for the 5th hydrological event

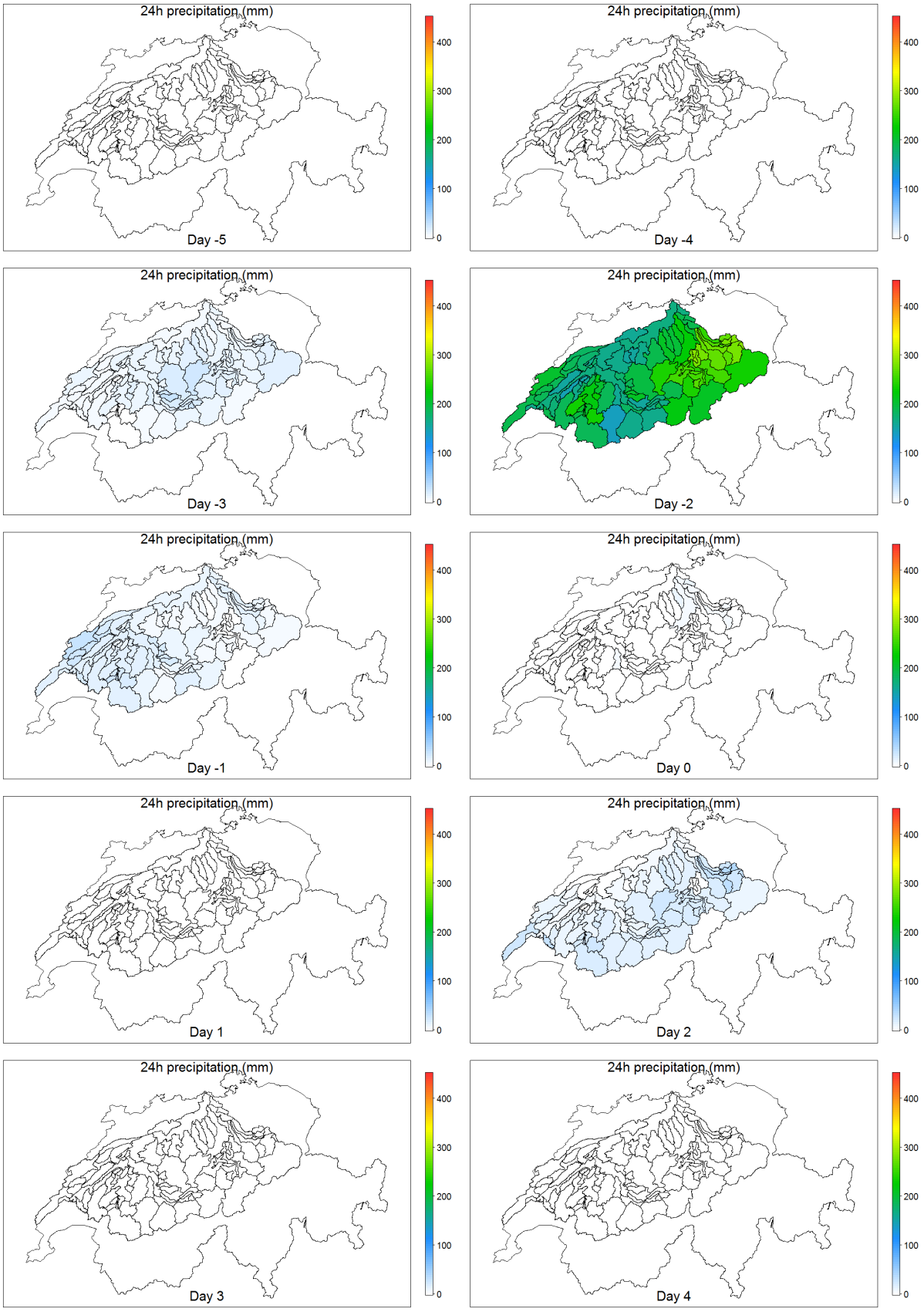


Figure A. 83 Space/time dynamics of precipitation during 10 days for the 5th hydrological event

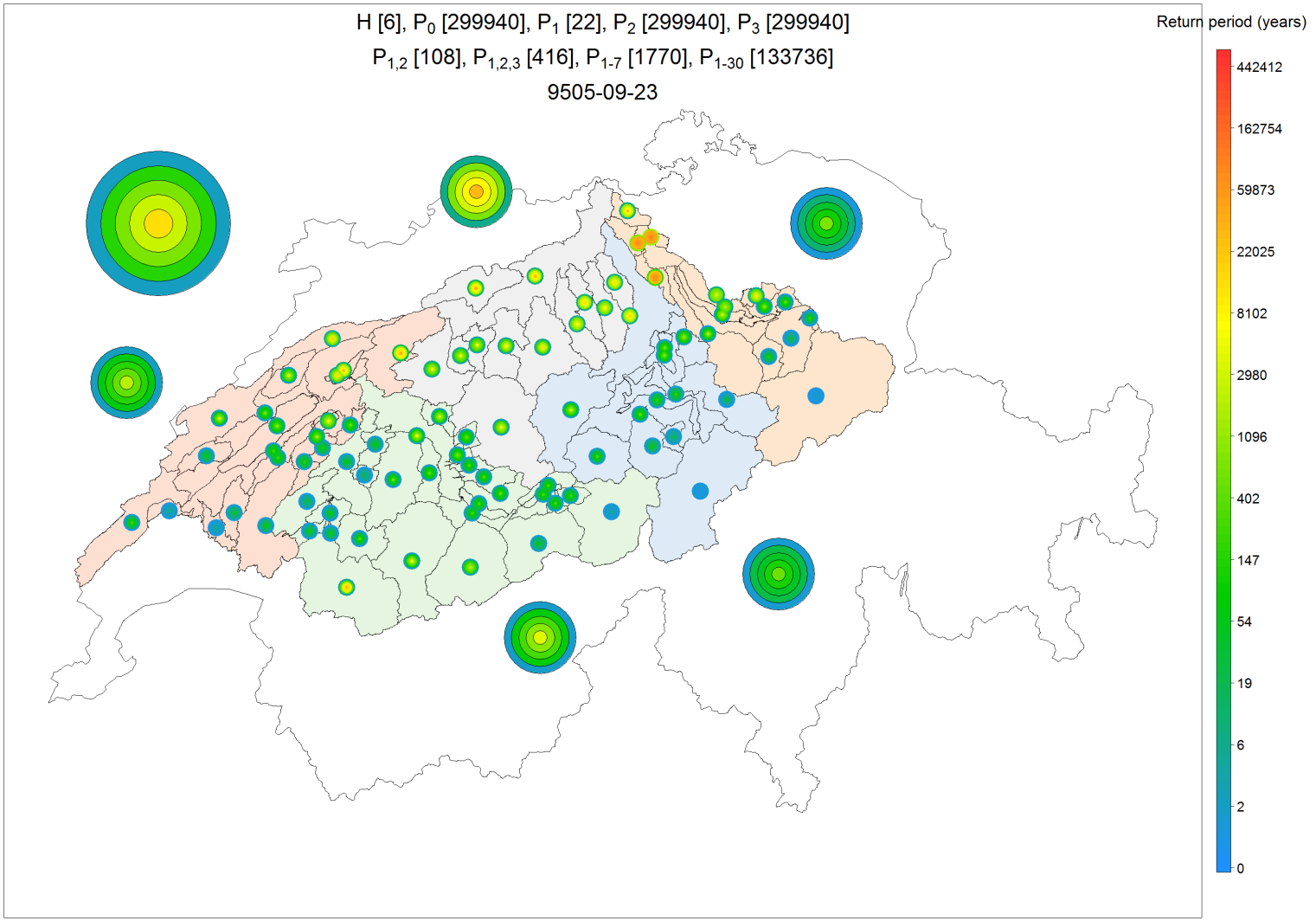


Figure A. 84 Severity map representing the severity at several spatial scales and for several accumulation periods for the 6thhydrological event

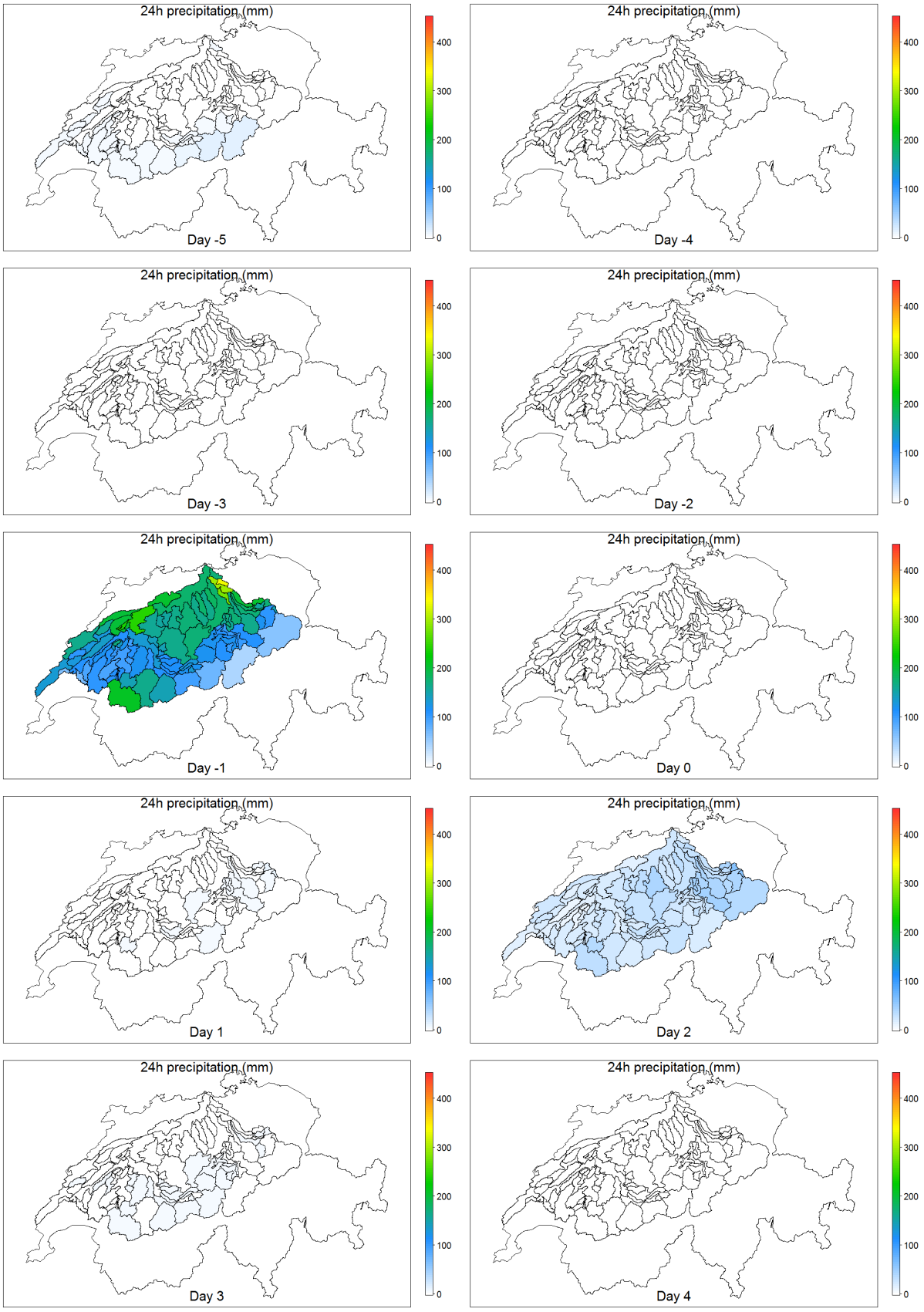


Figure A. 85 Space/time dynamics of precipitation during 10 days for the 6th hydrological event

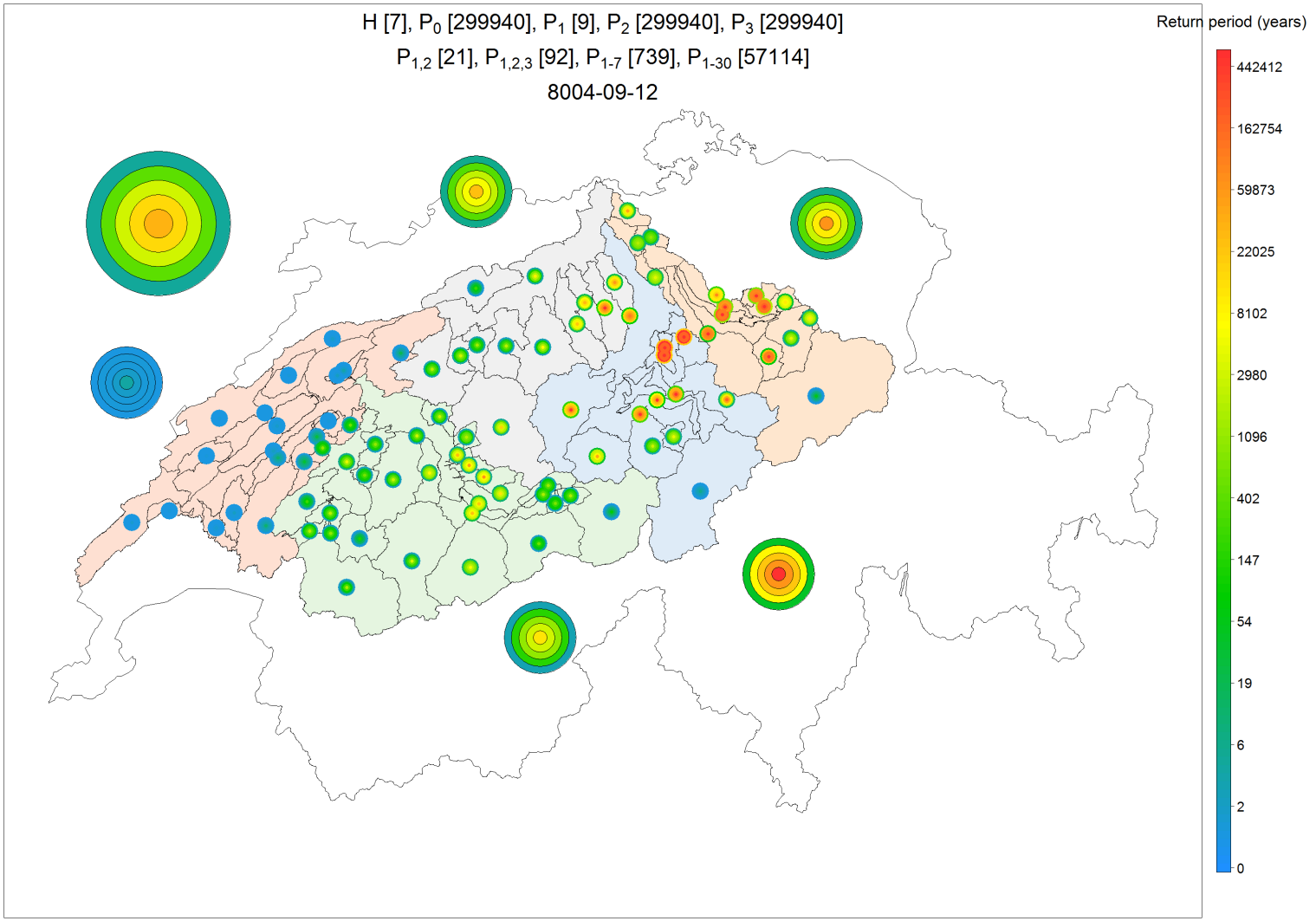


Figure A. 86 Severity map representing the severity at several spatial scales and for several accumulation periods for the 7th hydrological event

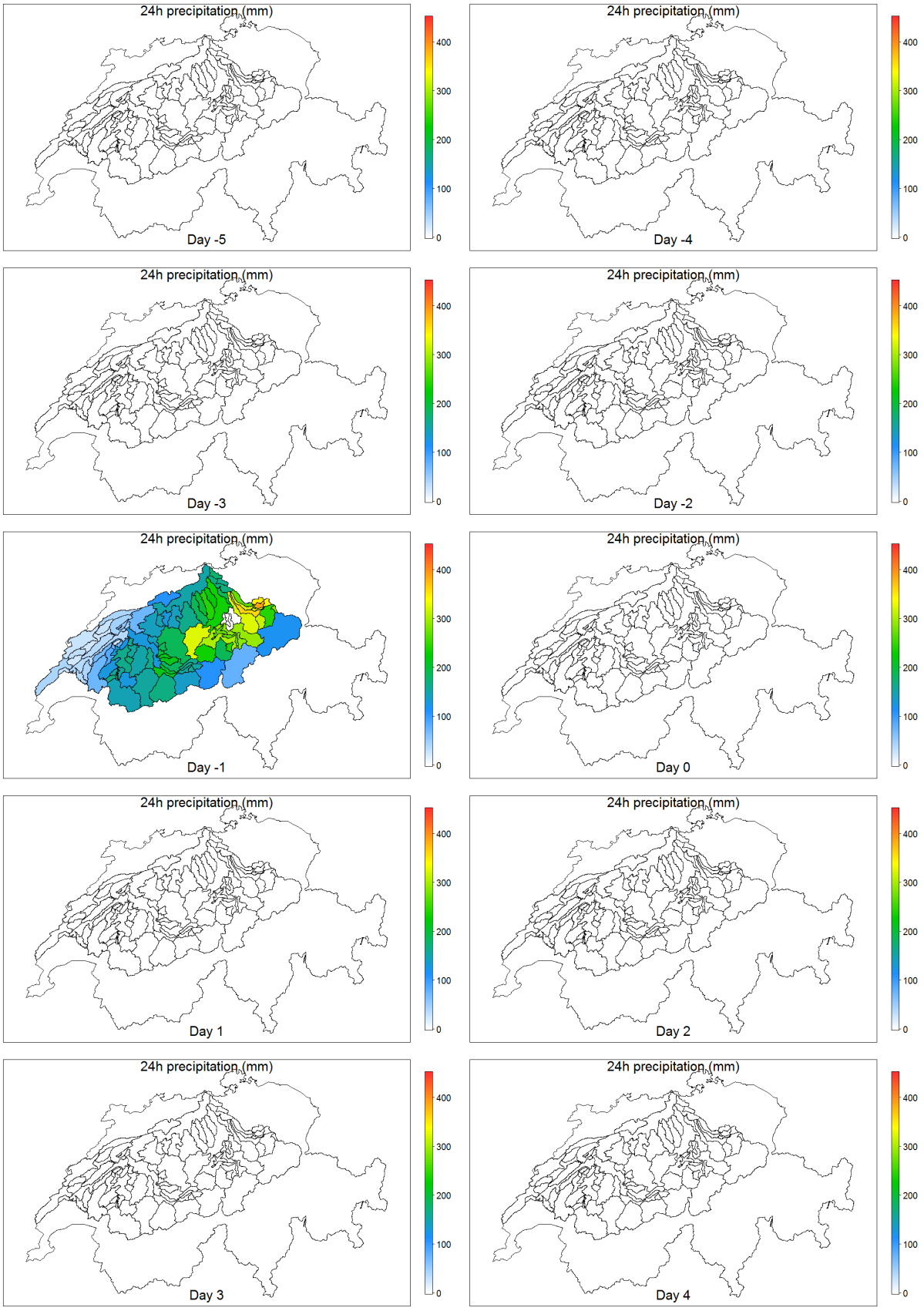


Figure A. 87 Space/time dynamics of precipitation during 10 days for the 7thhydrological event

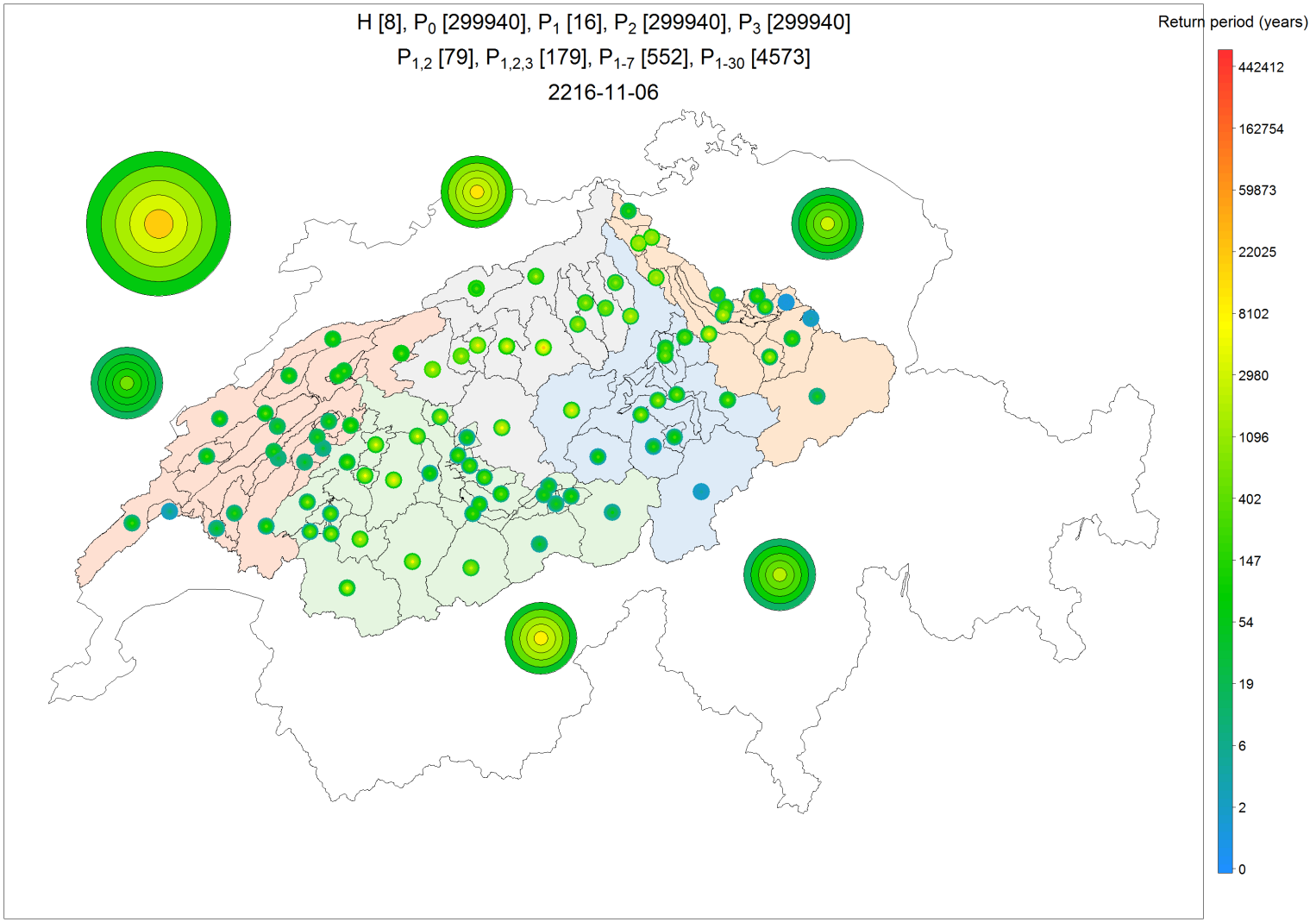


Figure A. 88 Severity map representing the severity at several spatial scales and for several accumulation periods for the 8th hydrological event

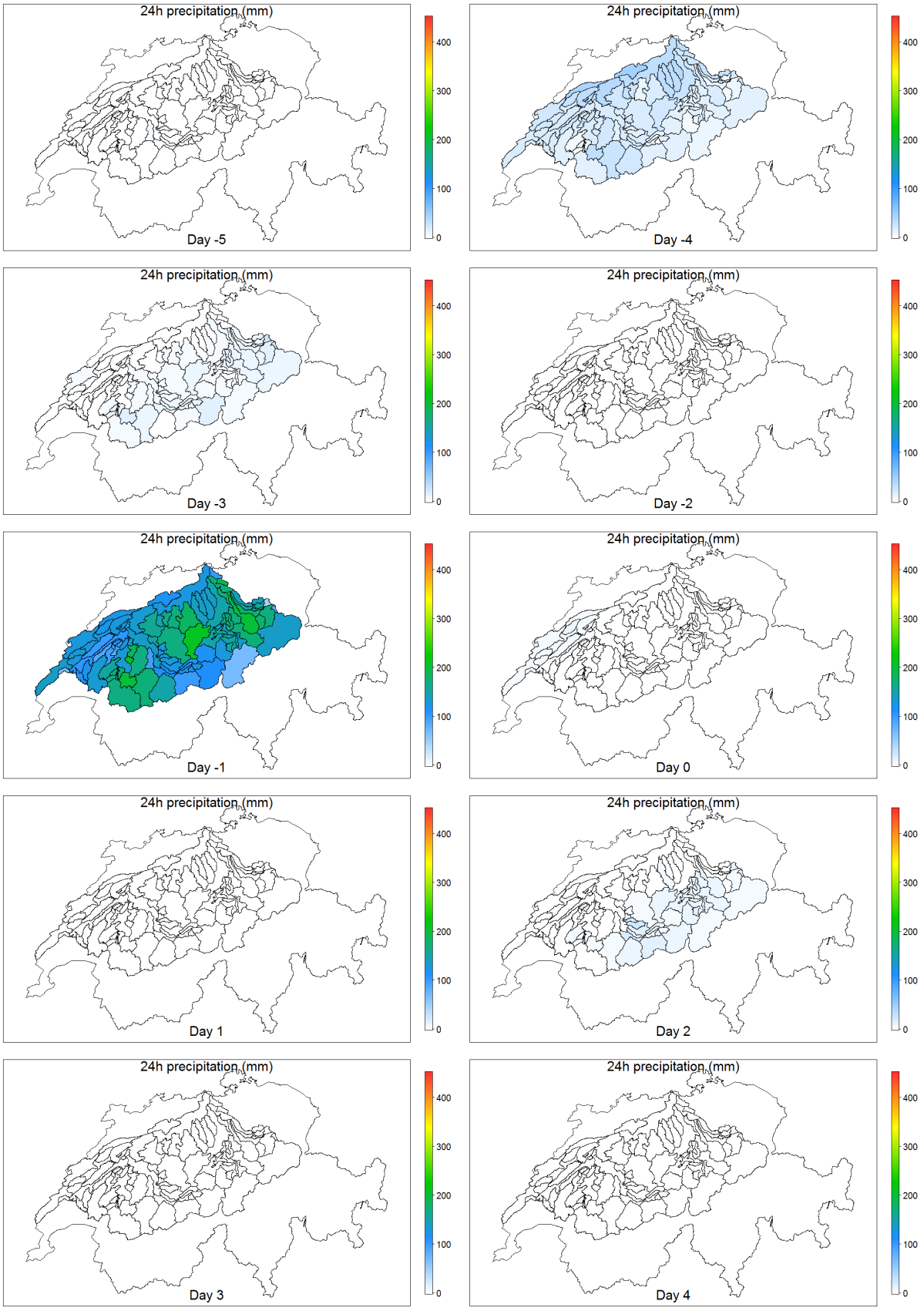


Figure A. 89 Space/time dynamics of precipitation during 10 days for the 8thhydrological event.

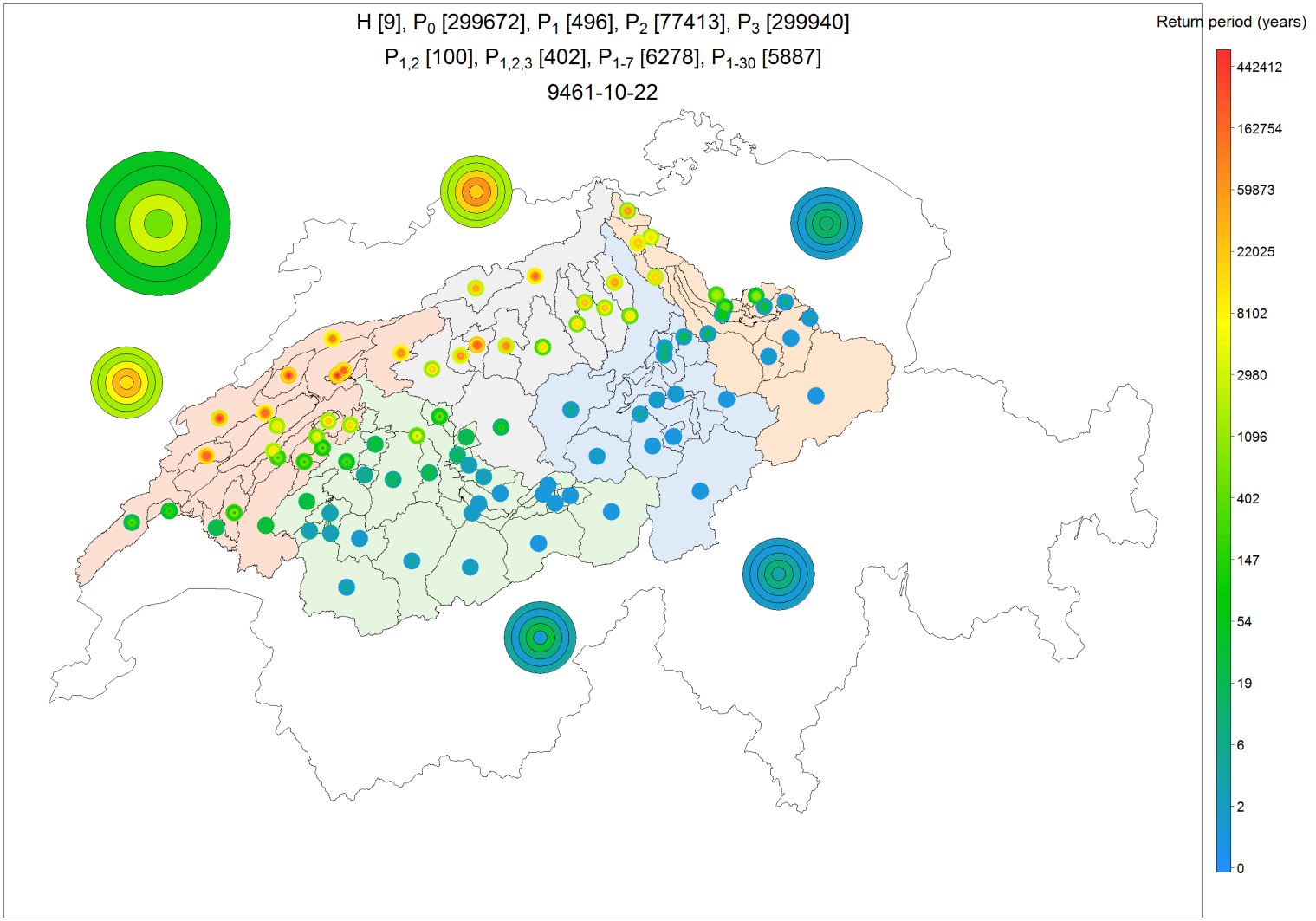


Figure A. 90 Severity map representing the severity at several spatial scales and for several accumulation periods for the 9thhydrological event

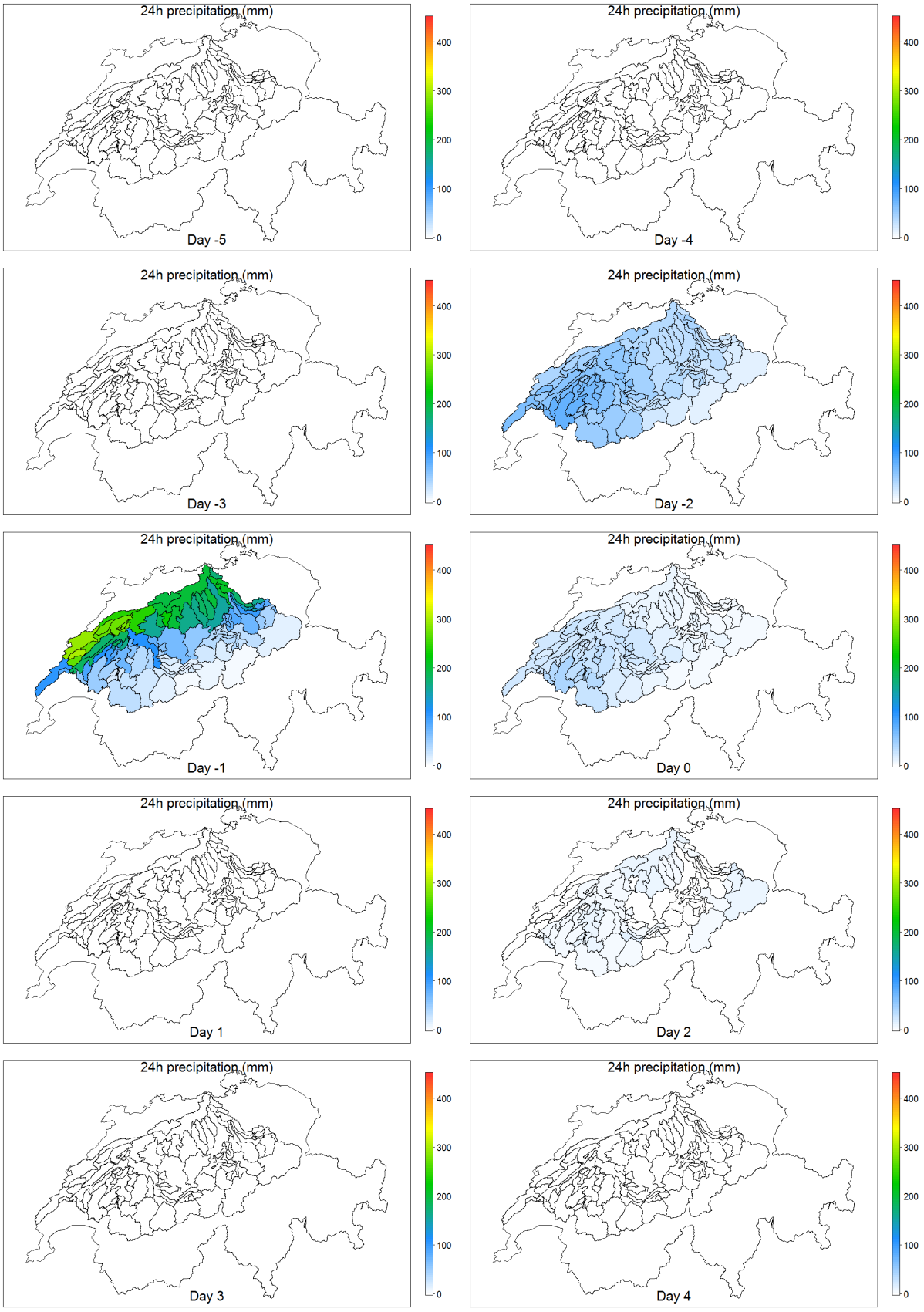


Figure A. 91 Space/time dynamics of precipitation during 10 days for the 9thhydrological event

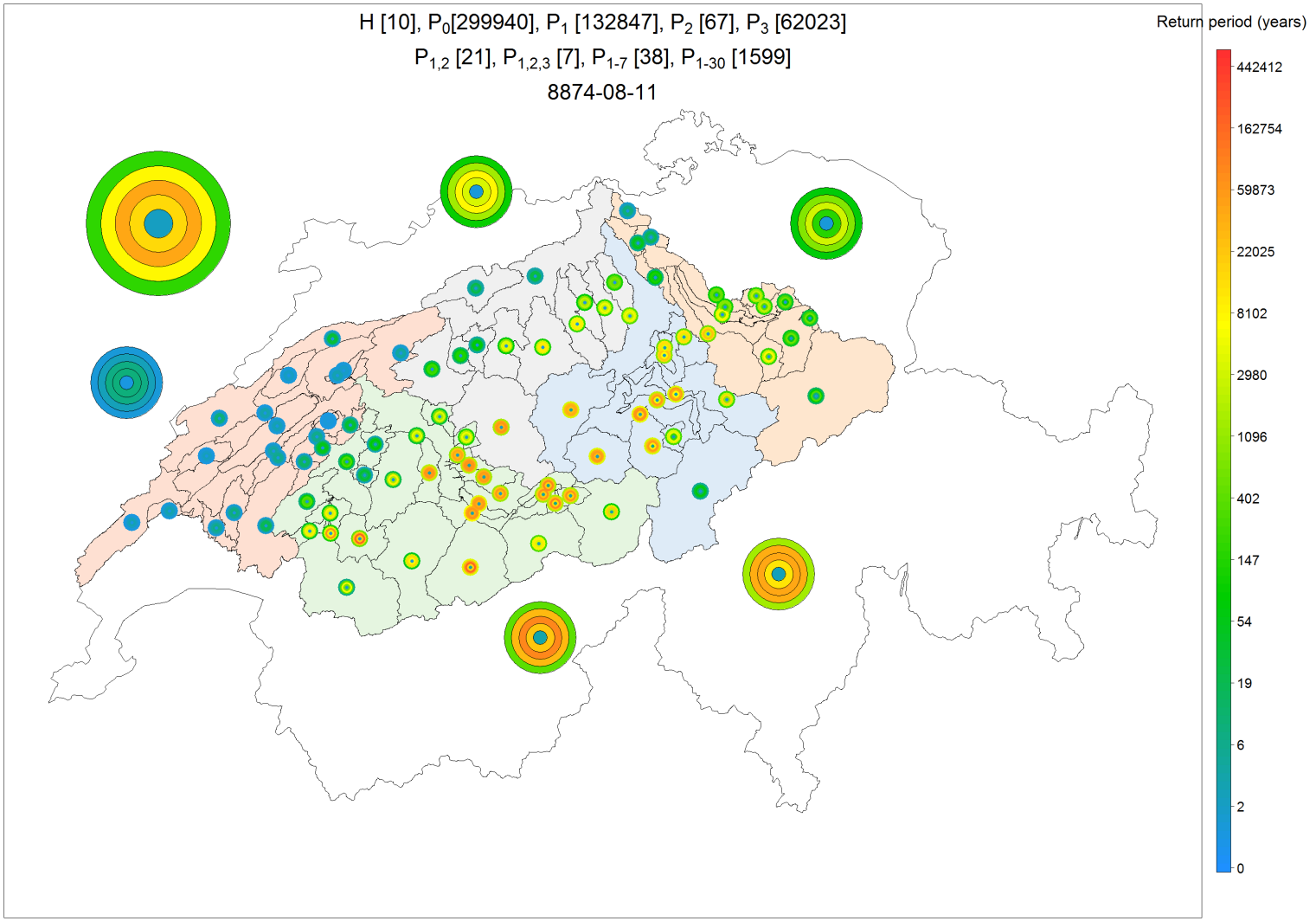


Figure A. 92 Severity map representing the severity at several spatial scales and for several accumulation periods for the 10th 1st hydrological event

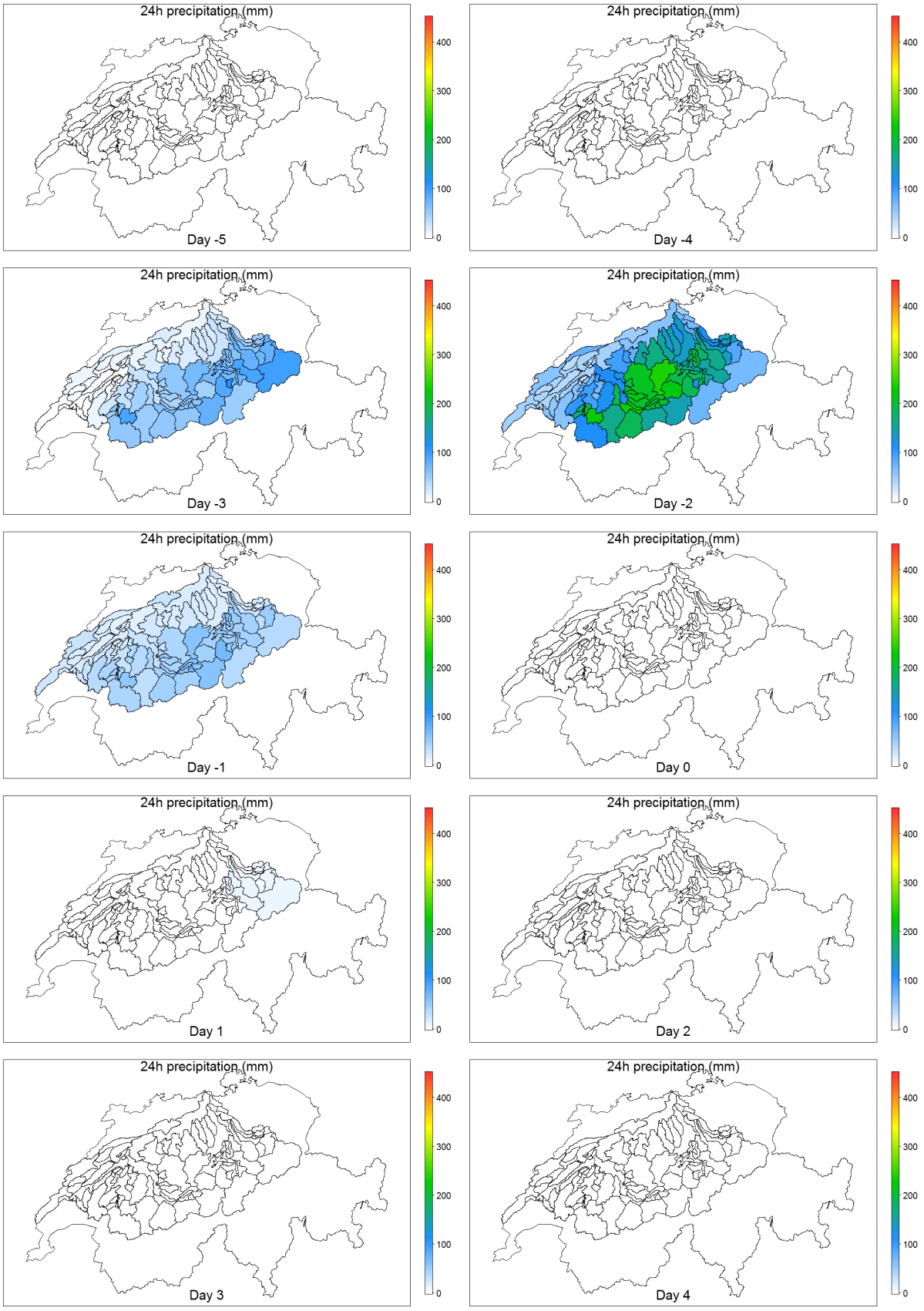


Figure A. 93 Space/time dynamics of precipitation during 10 days for the 10th hydrological event

10.5.5 Largest hydrological events with SCAMP: Severity maps and space/time dynamics

The precipitation space-time patterns during the largest hydrological events are shown in the following figures: for each event first the severity map is shown, then the space/time dynamics of precipitation. The severity map represents the severity of this event at several spatial scales and for several accumulation periods; the space/time dynamics of precipitation for 10 days are composed of 10 maps ("Day 0" is the day of the discharge peak).

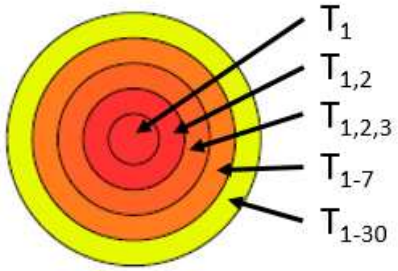
The title of the severity map shows the rank (in descending order) of the precipitation events at the whole catchment scale for several accumulation durations:

- H[]: rank of the hydrological event
- P_0[]: rank of the 1-day precipitation during the day of the hydrological peak (“Day 0”)
- P_1[]: rank of the 1-day precipitation during the day before the hydrological peak (“Day -1”)
- P_2[]: rank of the 1-day precipitation 2 days before the hydrological peak (“Day -2”)
- P_3[]: rank of the 1-day precipitation 3 days before the hydrological peak (“Day -3”)
- $P_{1,2}$[]: rank of the 2-day precipitation from “Day -1” to “Day -2”
- $P_{1,2,3}$[]: rank of the 3-day precipitation from “Day -1” to “Day -3”
- $P_{1\text{-}7}$[]: rank of the 7-day precipitation from “Day -1” to “Day -7”
- $P_{1\text{-}30}$[]: rank of the 30-day precipitation from “Day -1” to “Day -30”
- In addition, the fictive date of the hydrological peak is indicated on the third line

The severity of the simulated event is then represented with a set of severity-board. Each severity-board gives the return period of precipitation simulated for this event when cumulated over several durations (all ending the day before the hydrological peak):

- 1 day (center of the severity-board)
- 2 days
- 3 days
- 7 days
- 30 days (edge of the severity-board)

For instance, if T_1, $T_{1,2}$, $T_{1,2,3}$, $T_{1\text{-}7}$, and $T_{1\text{-}30}$ are the return periods associated with the precipitation P_1, $P_{1,2}$, $P_{1,2,3}$, $P_{1\text{-}7}$, and $P_{1\text{-}30}$, respectively



The color scale for the severity-boards is given in the right-hand side of the figure. Severity-boards are given for precipitation simulated over different spatial scales:

- local precipitation at each of the 89 simulation sub-catchments (small circles)
- mean areal precipitation over each of the 5 main sub-catchments (medium circles)
- whole catchment (large circle)

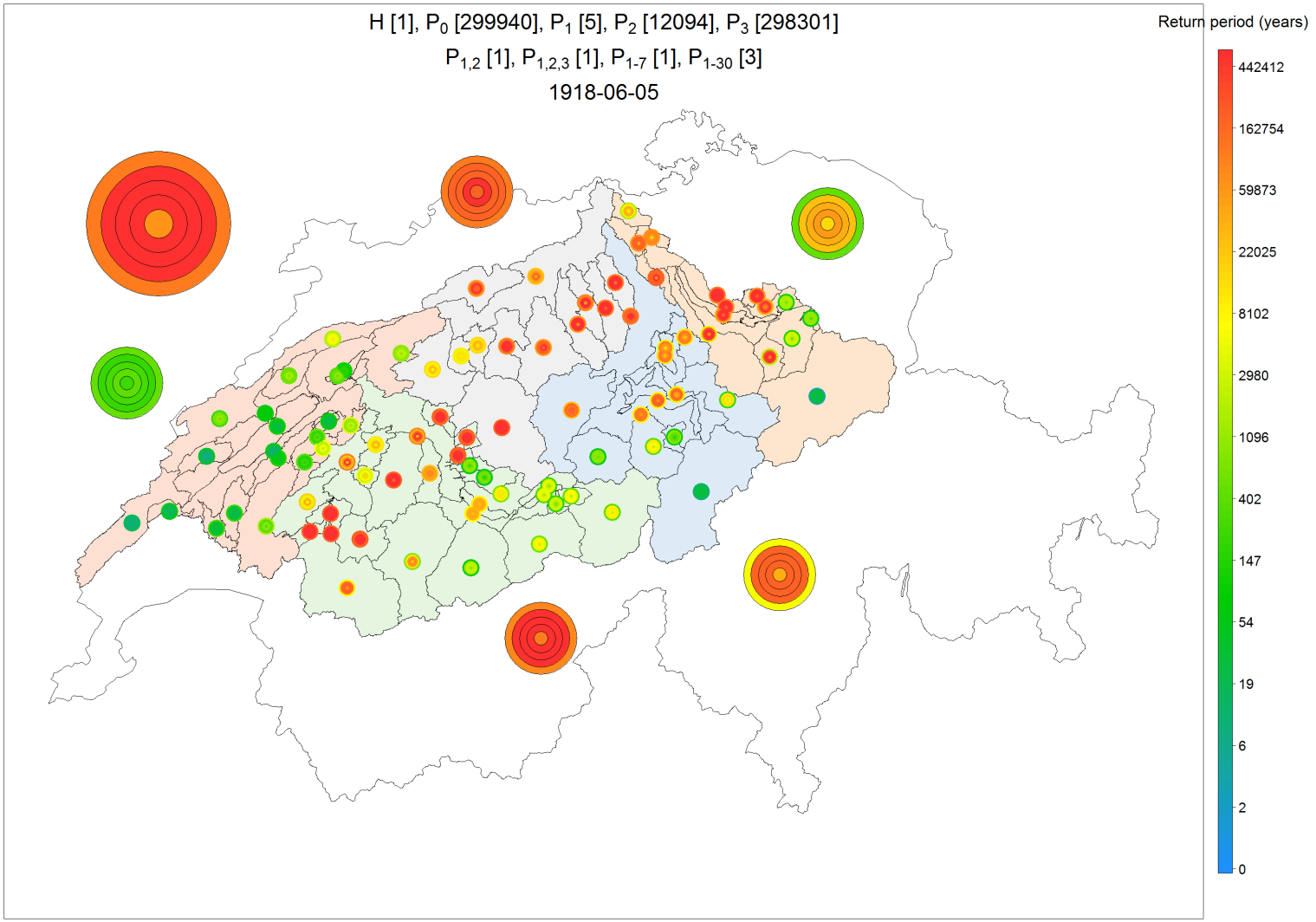


Figure A. 94 Severity map representing the severity at several spatial scales and for several accumulation periods for the 1st hydrological event

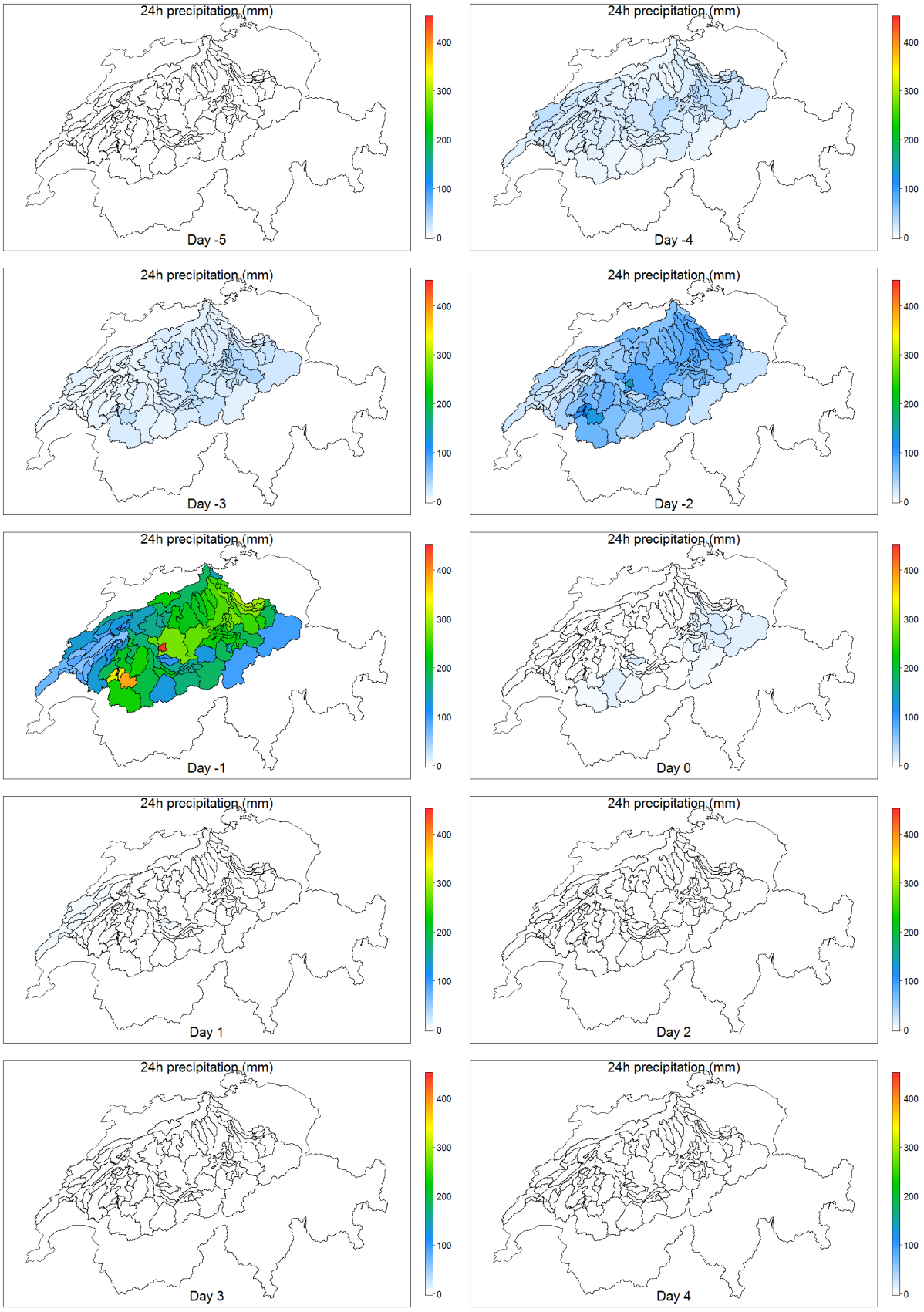


Figure A. 95 Space/time dynamics of precipitation during 10 days for the1st hydrological event

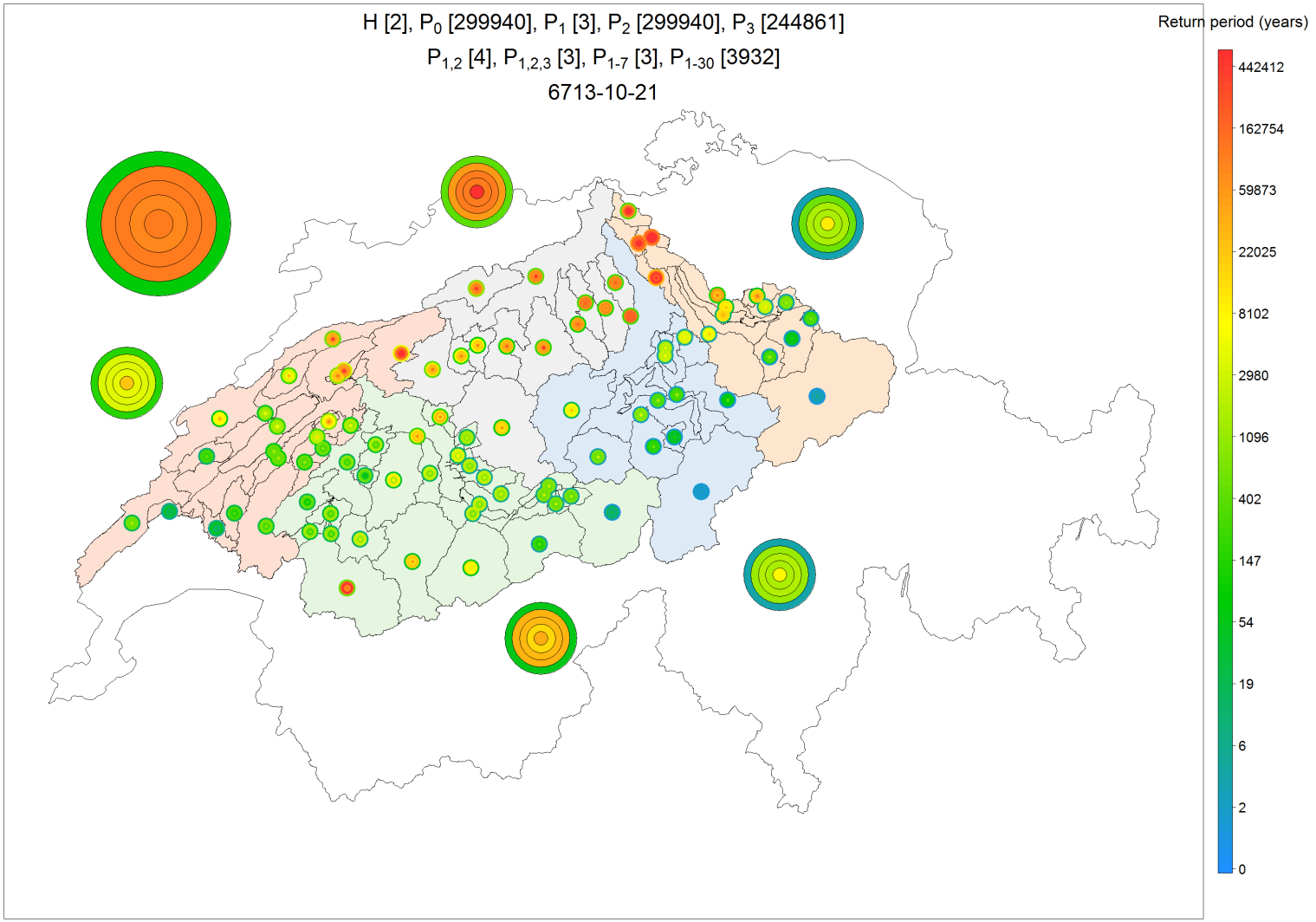


Figure A. 96 Severity map representing the severity at several spatial scales and for several accumulation periods for the 2nd hydrological event

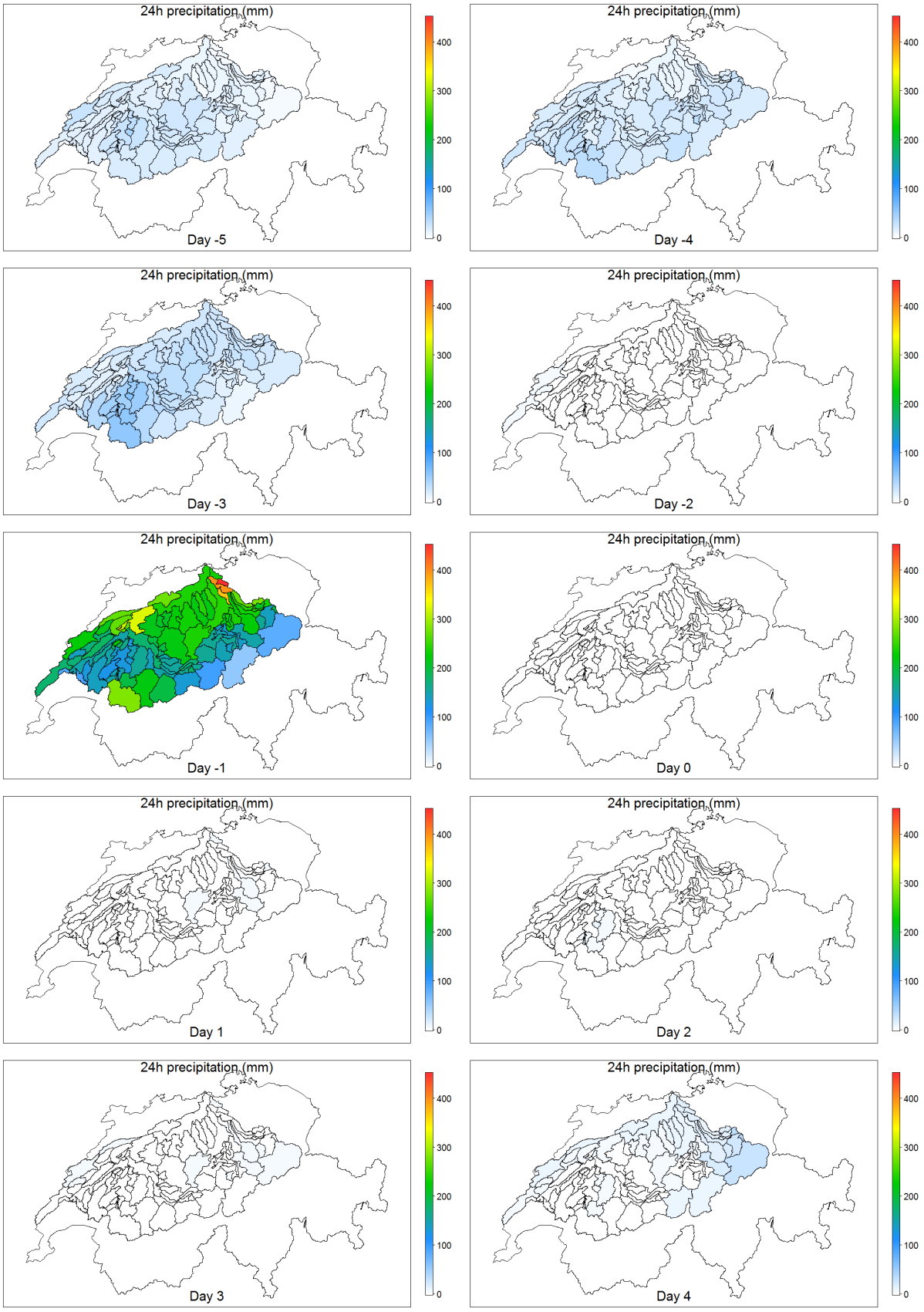


Figure A. 97 Space/time dynamics of precipitation during 10 days for the 2nd hydrological event

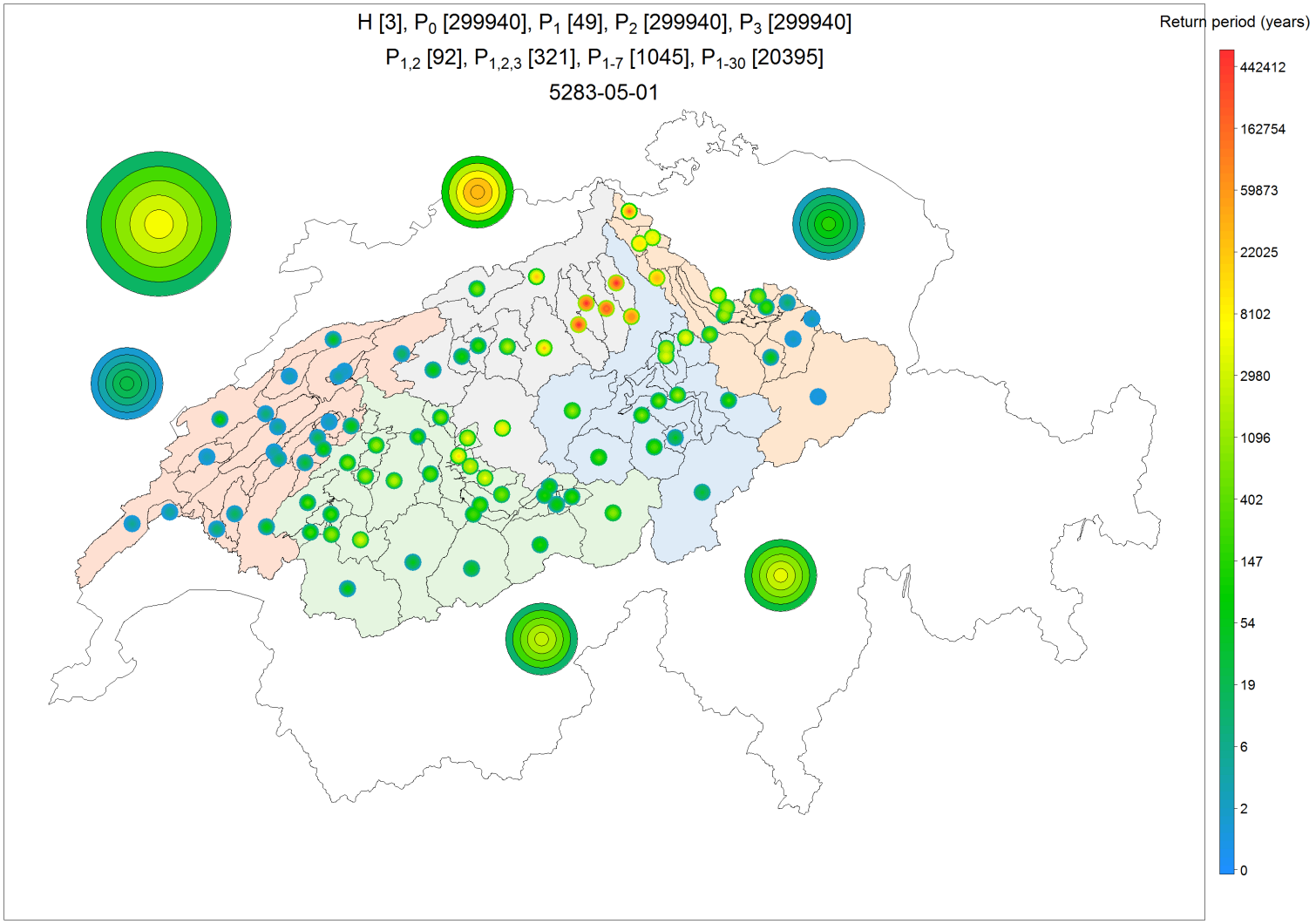


Figure A. 98 Severity map representing the severity at several spatial scales and for several accumulation periods for the 3rd hydrological event

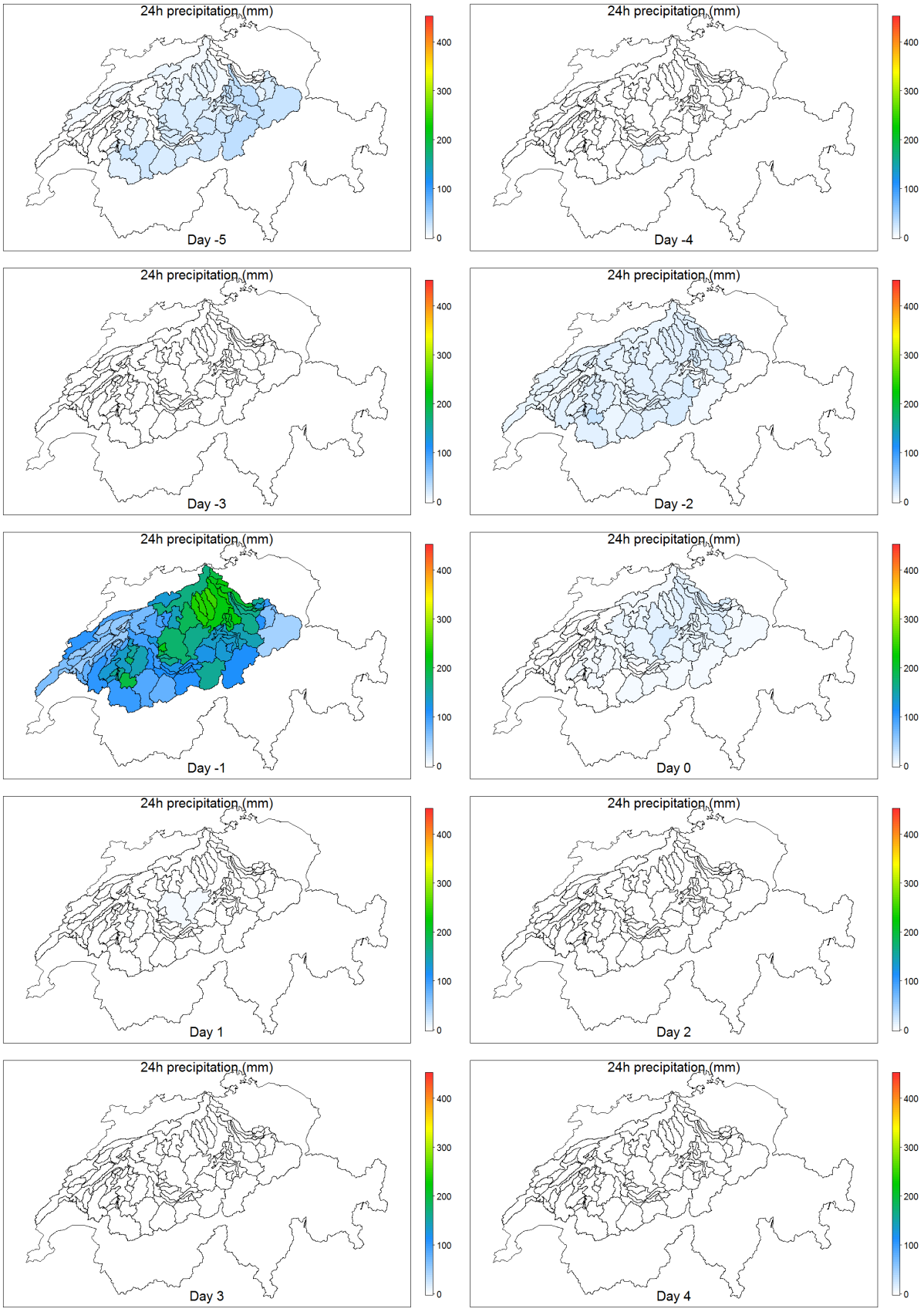


Figure A. 99 Space/time dynamics of precipitation during 10 days for the 3rdhydrological event

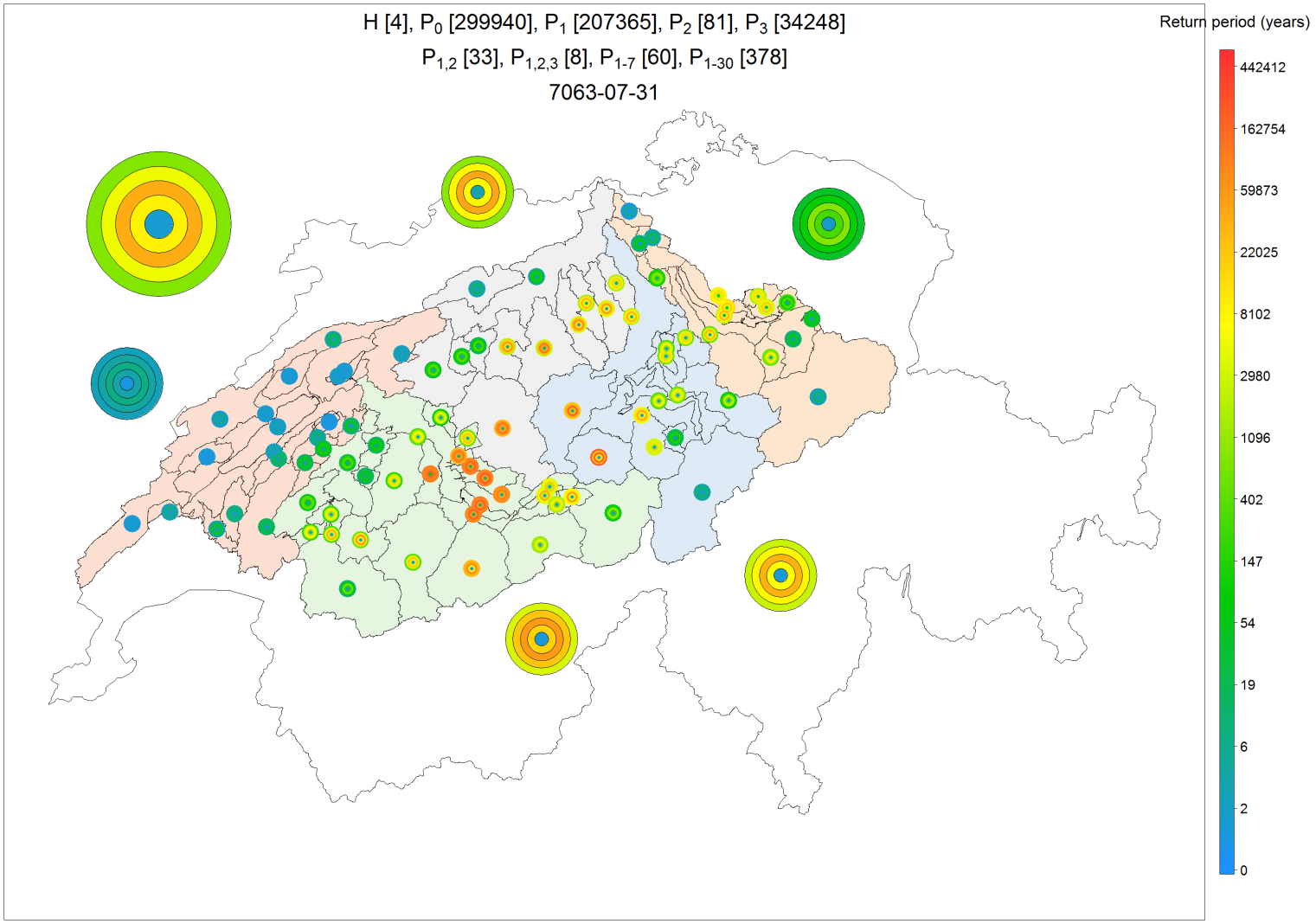


Figure A. 100 Severity map representing the severity at several spatial scales and for several accumulation periods for the 4th hydrological event

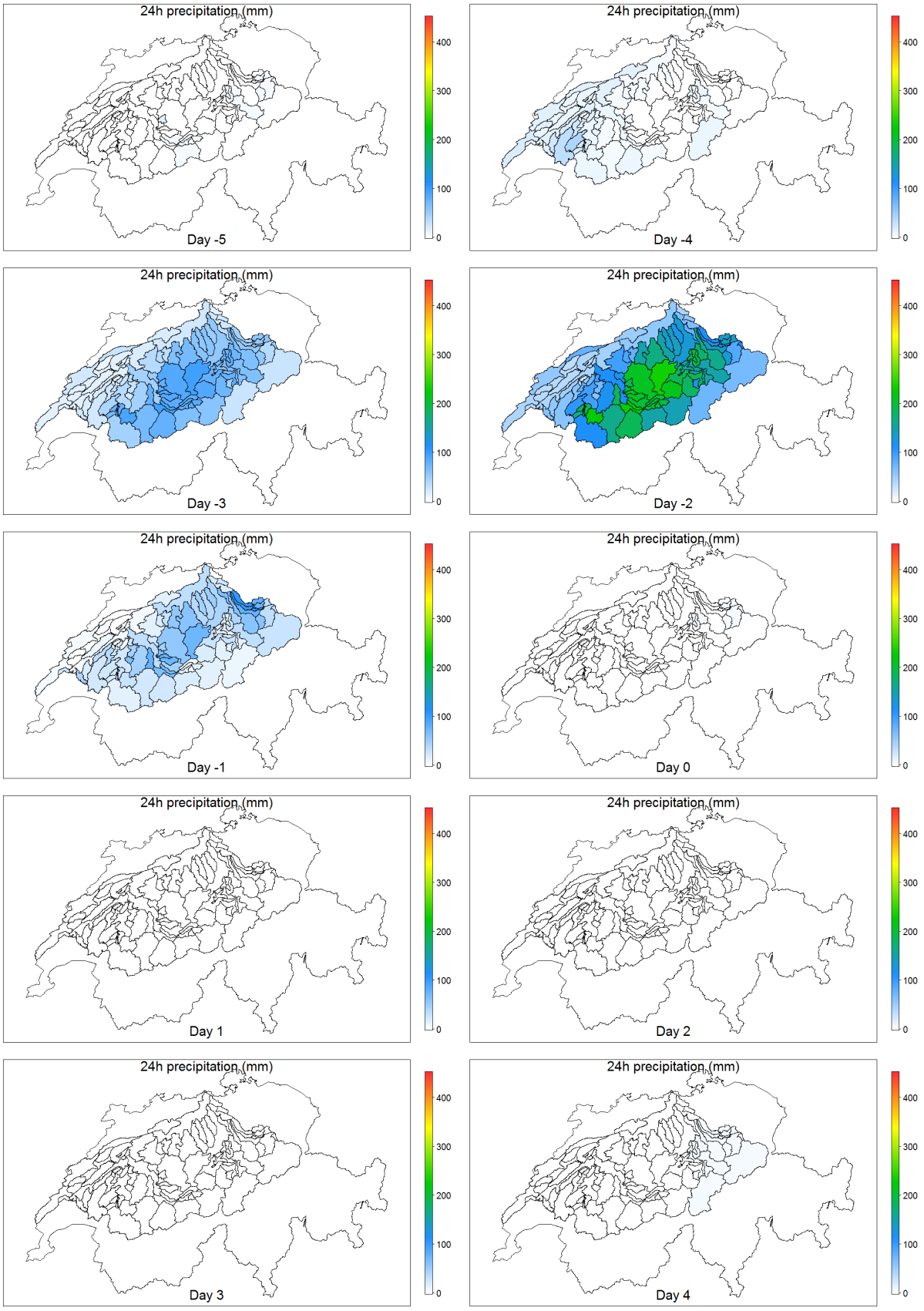


Figure A. 101 Space/time dynamics of precipitation during 10 days for the 4th t hydrological event

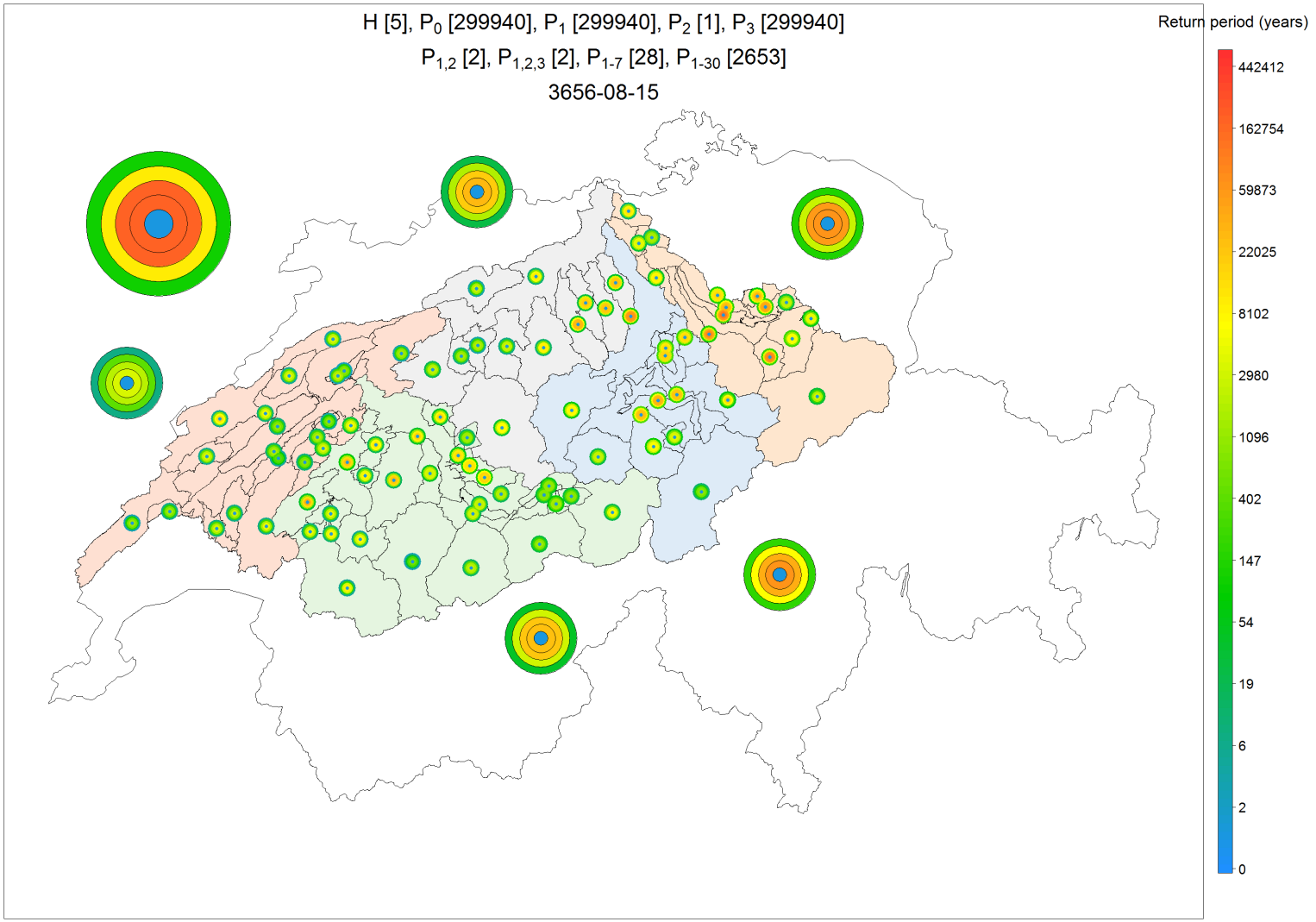


Figure A. 102 Severity map representing the severity at several spatial scales and for several accumulation periods for the 5th hydrological event

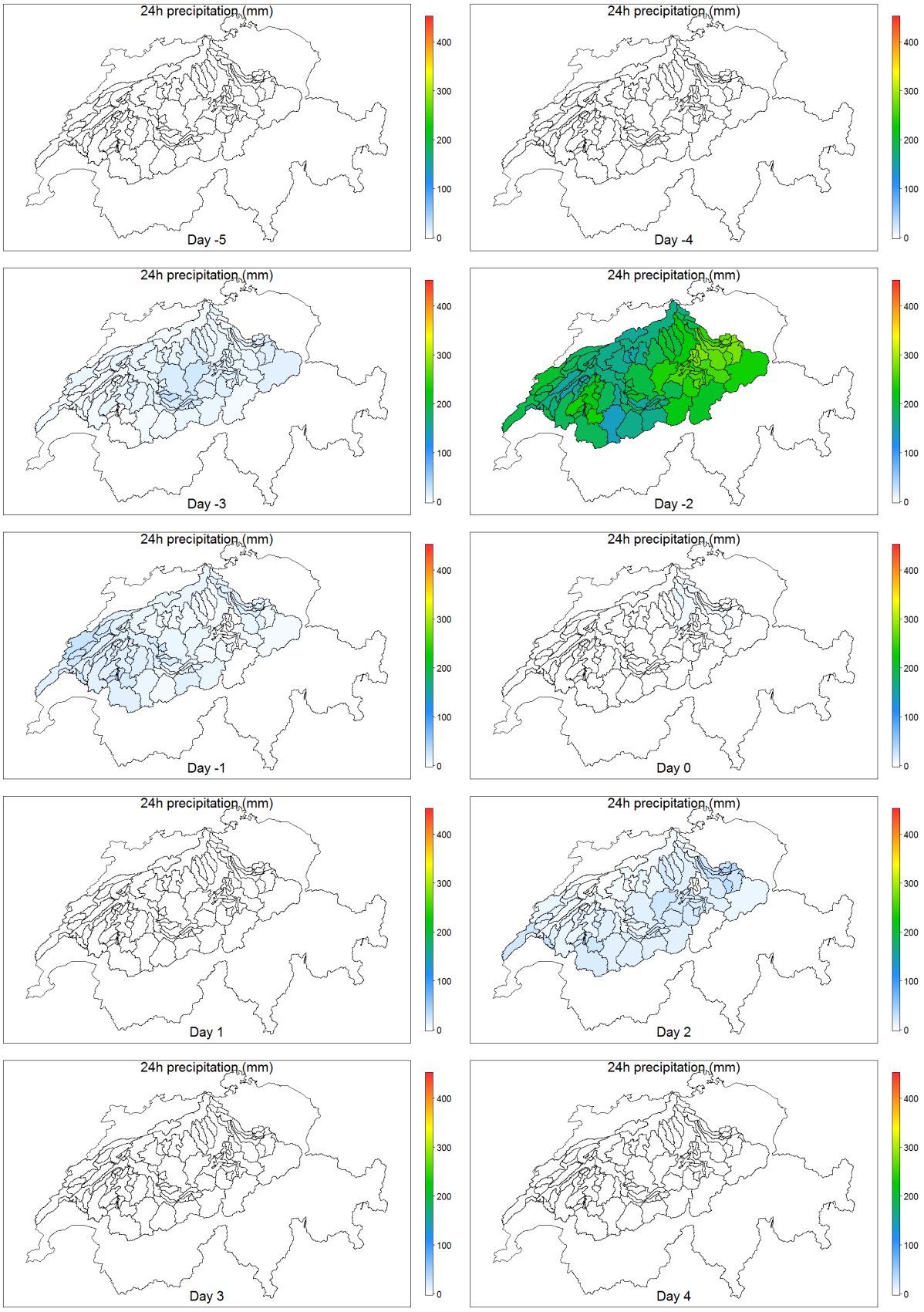


Figure A. 103 Space/time dynamics of precipitation during 10 days for the 5th hydrological event

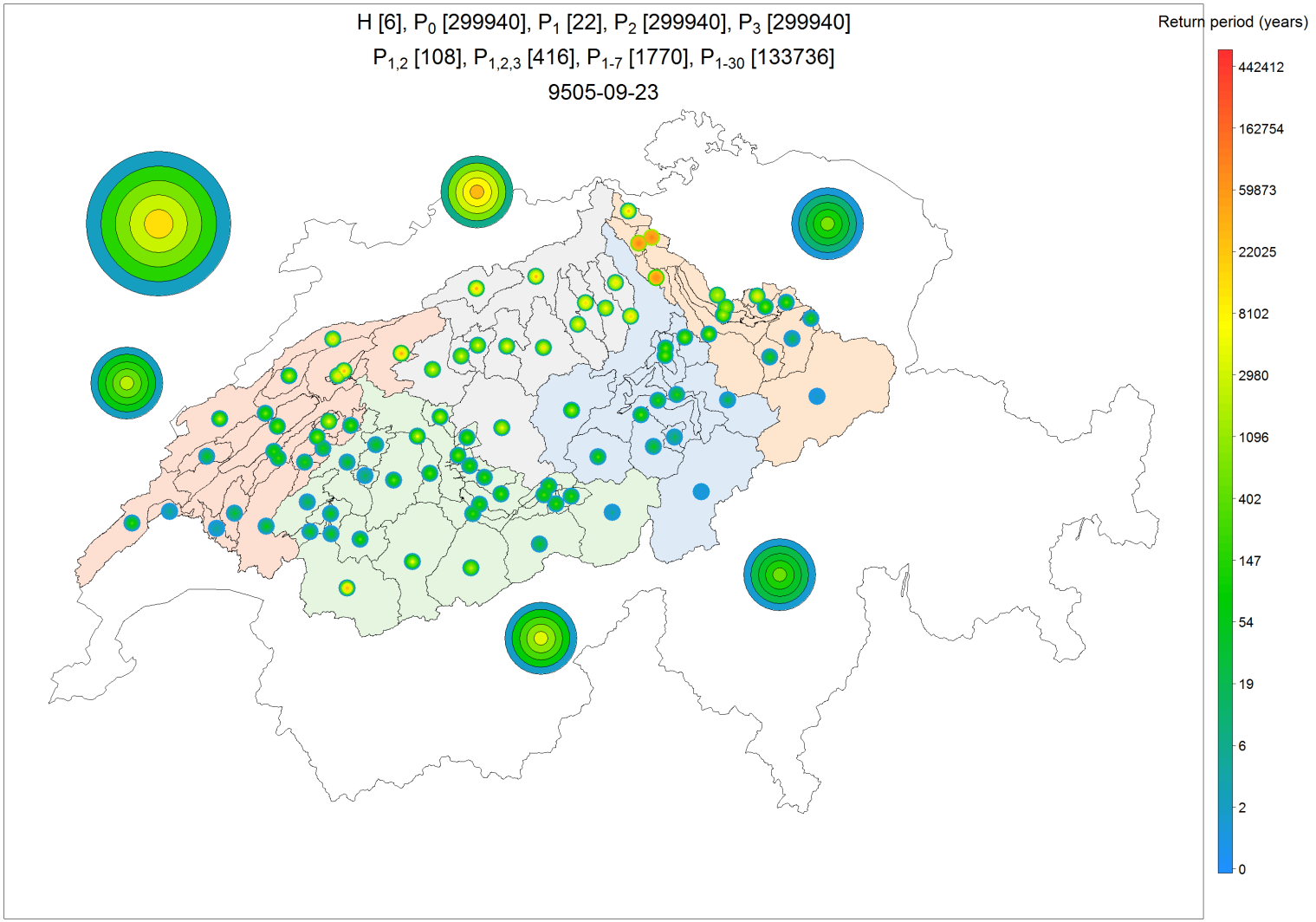


Figure A. 104 Severity map representing the severity at several spatial scales and for several accumulation periods for the 6thhydrological event

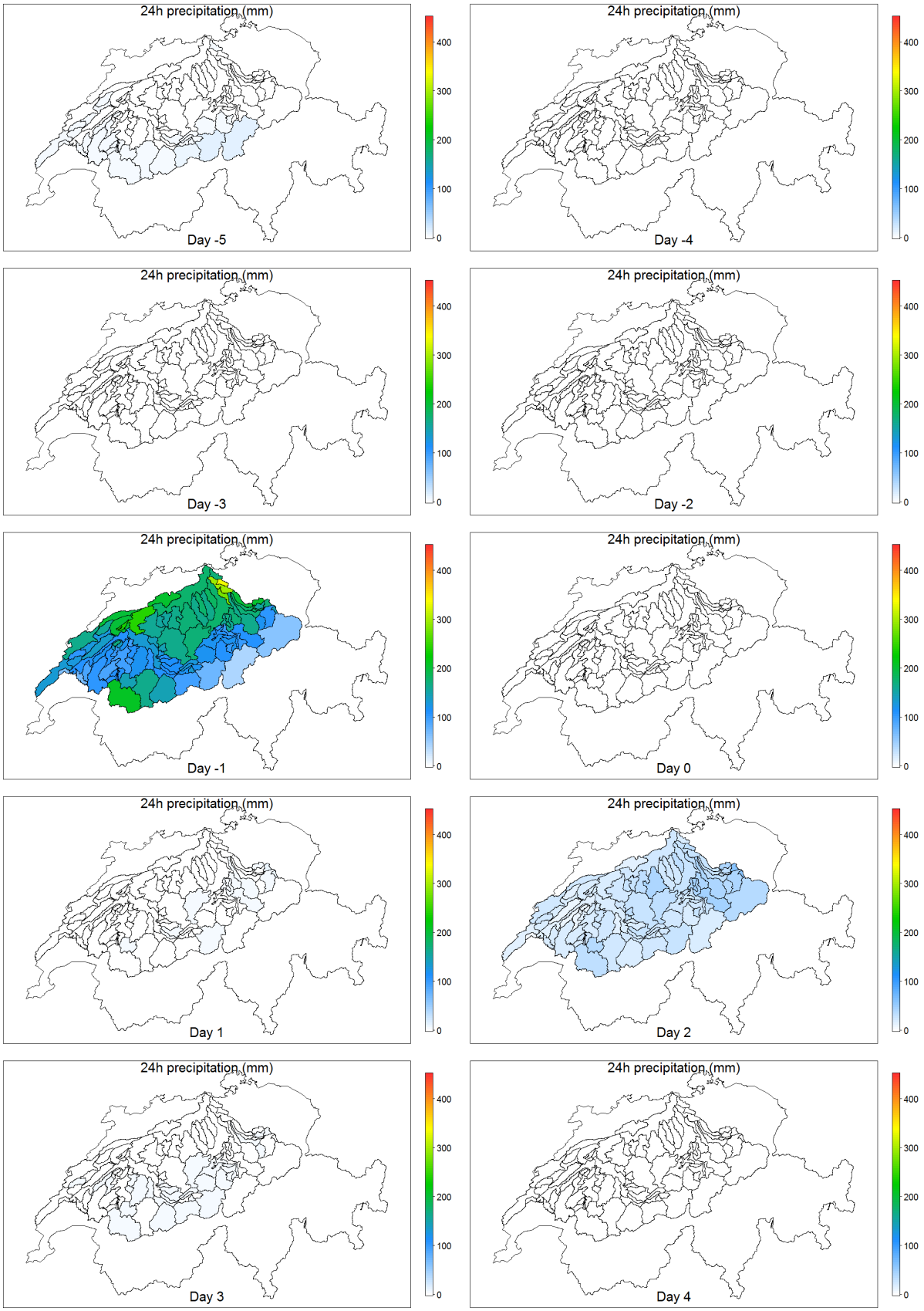


Figure A. 105 Space/time dynamics of precipitation during 10 days for the 6th hydrological event

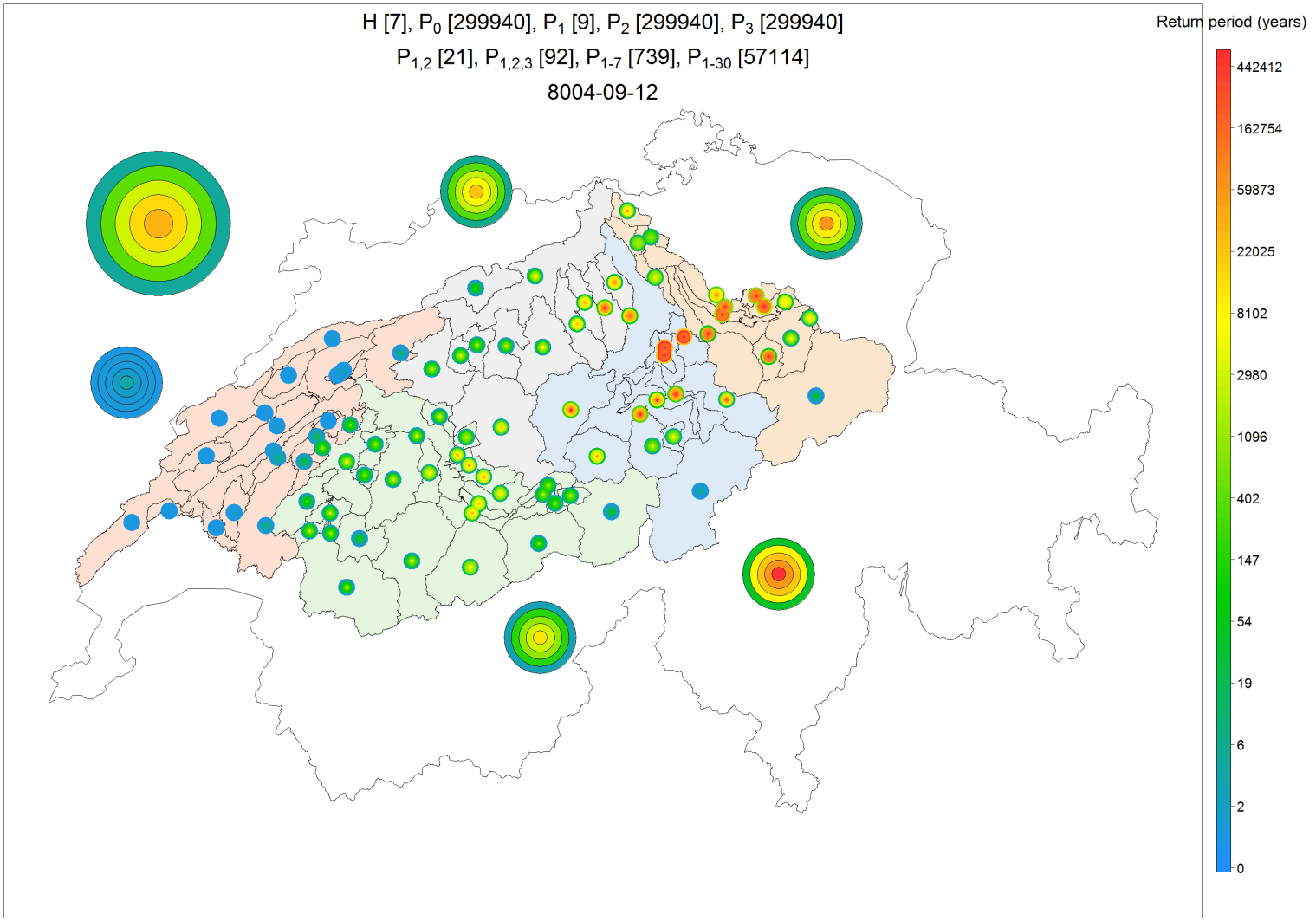


Figure A. 106 Severity map representing the severity at several spatial scales and for several accumulation periods for the 7th hydrological event

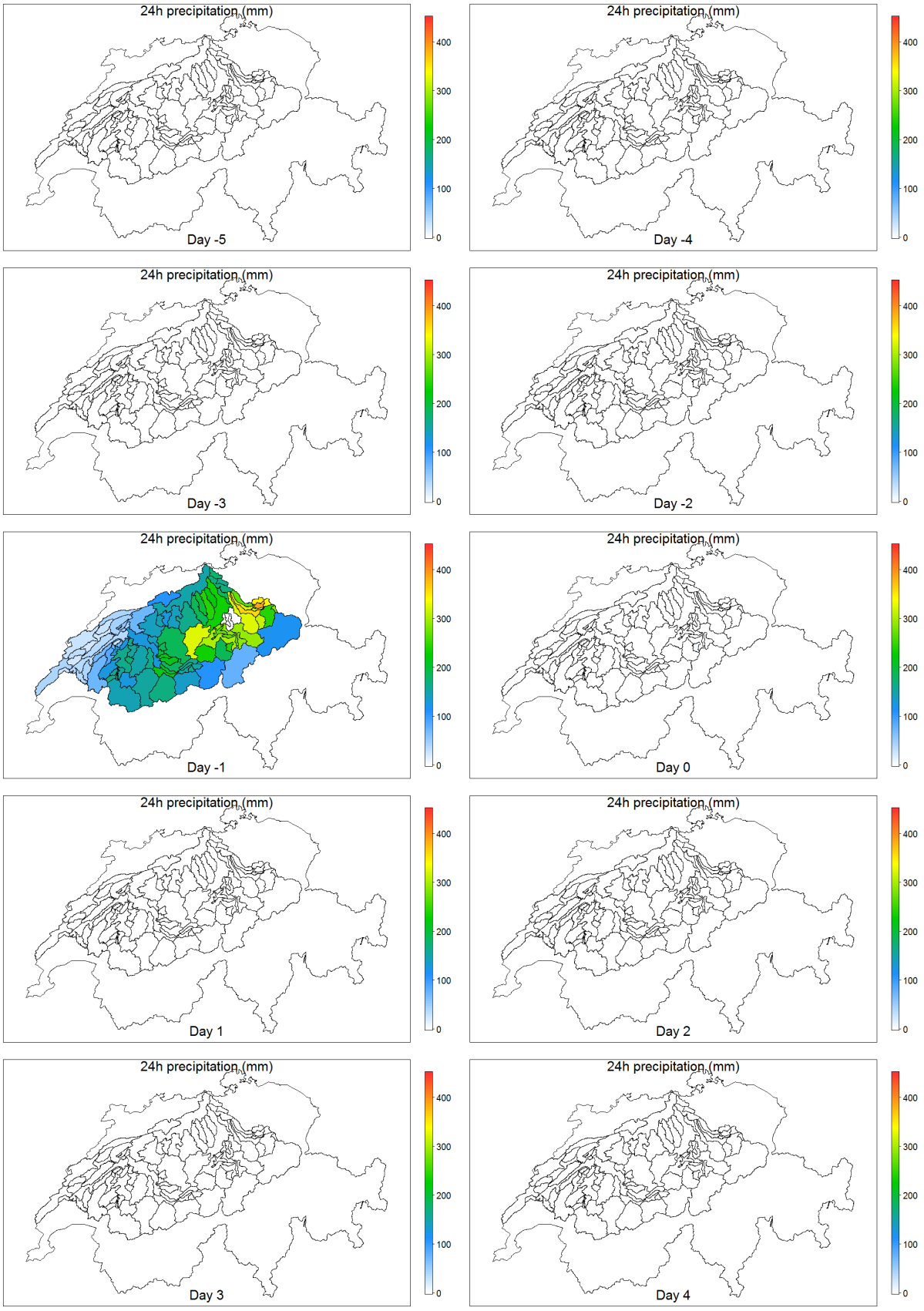


Figure A. 107 Space/time dynamics of precipitation during 10 days for the 7thhydrological event

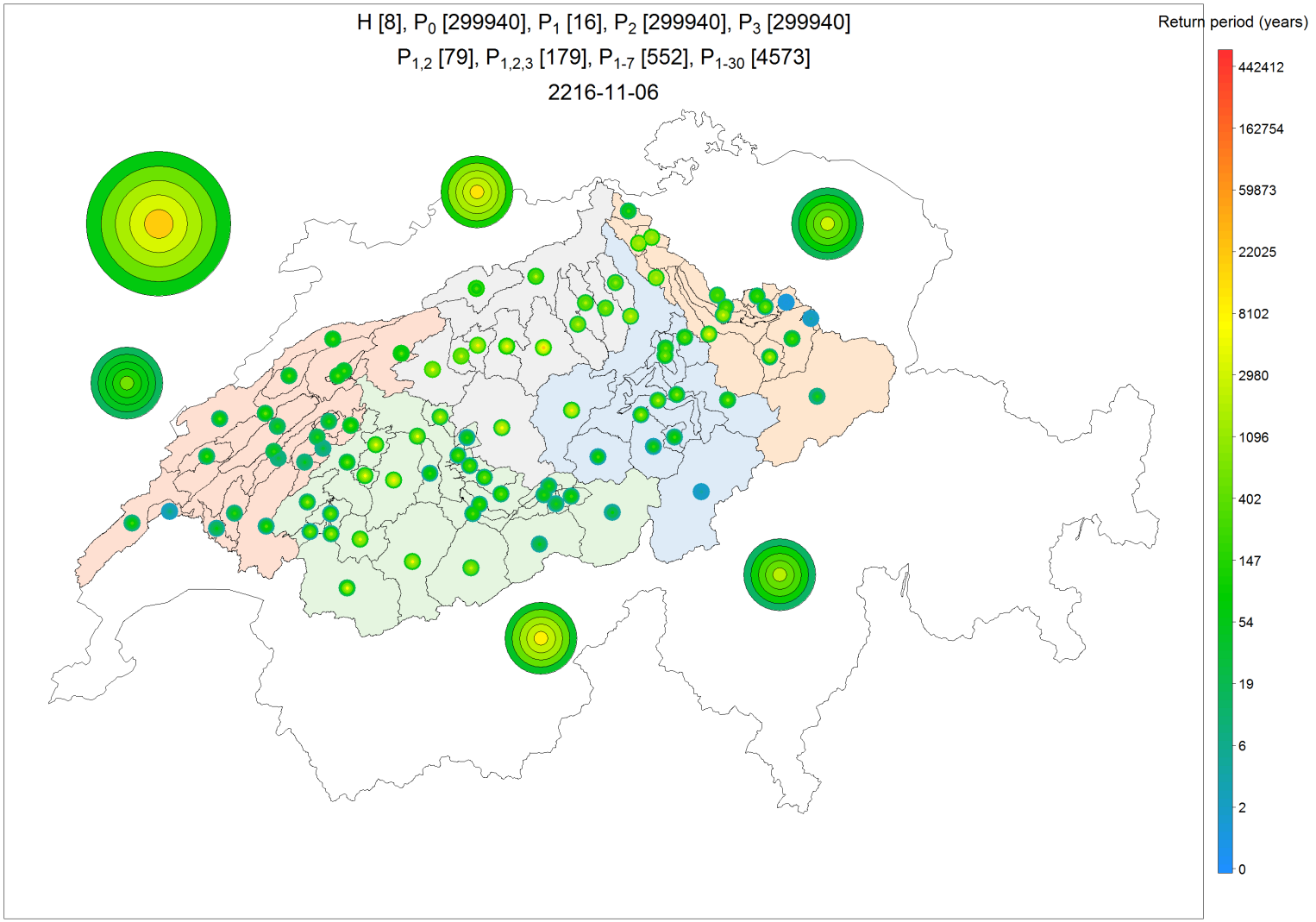


Figure A. 108 Severity map representing the severity at several spatial scales and for several accumulation periods for the 8th hydrological event

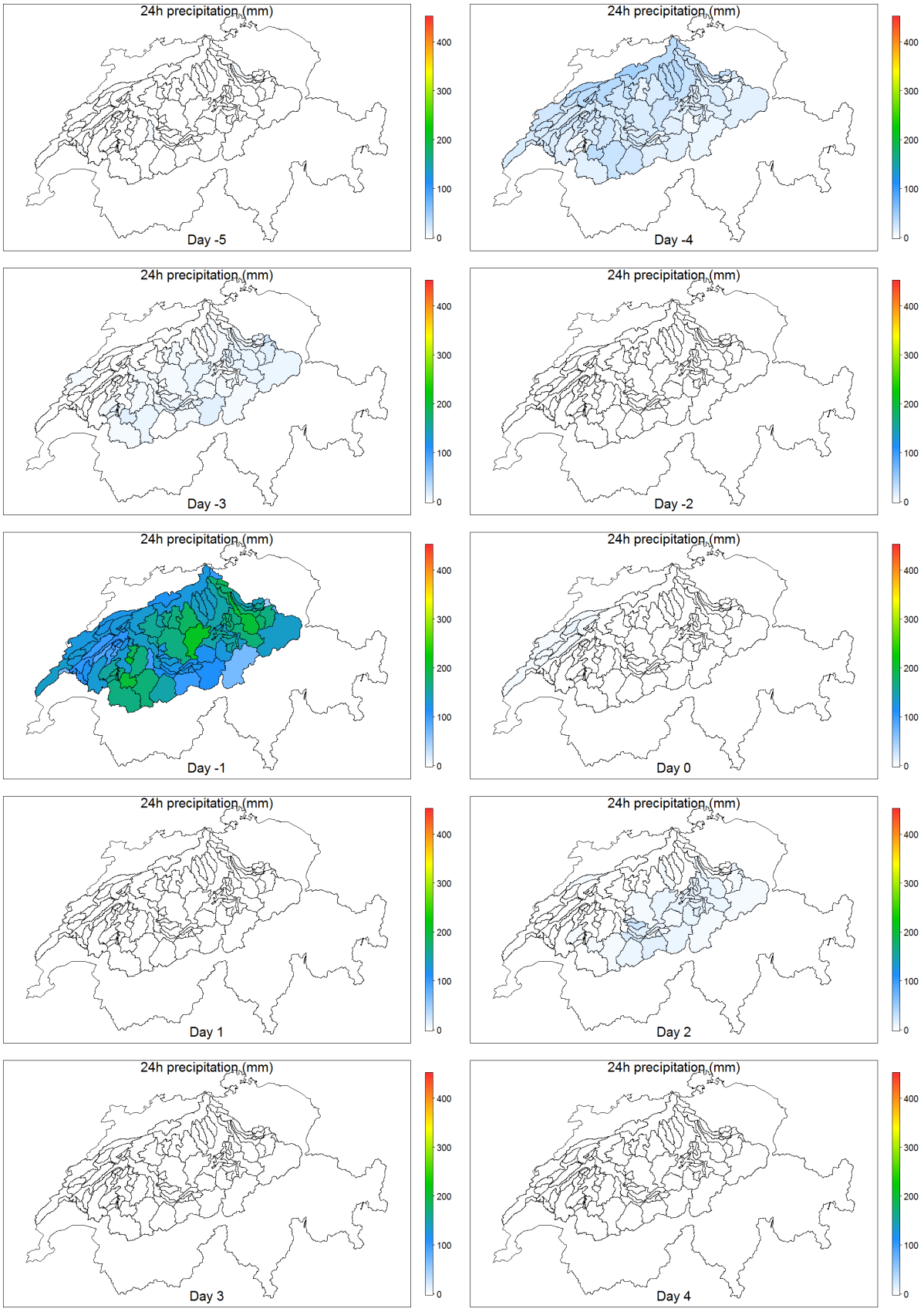


Figure A. 109 Space/time dynamics of precipitation during 10 days for the 8thhydrological event.

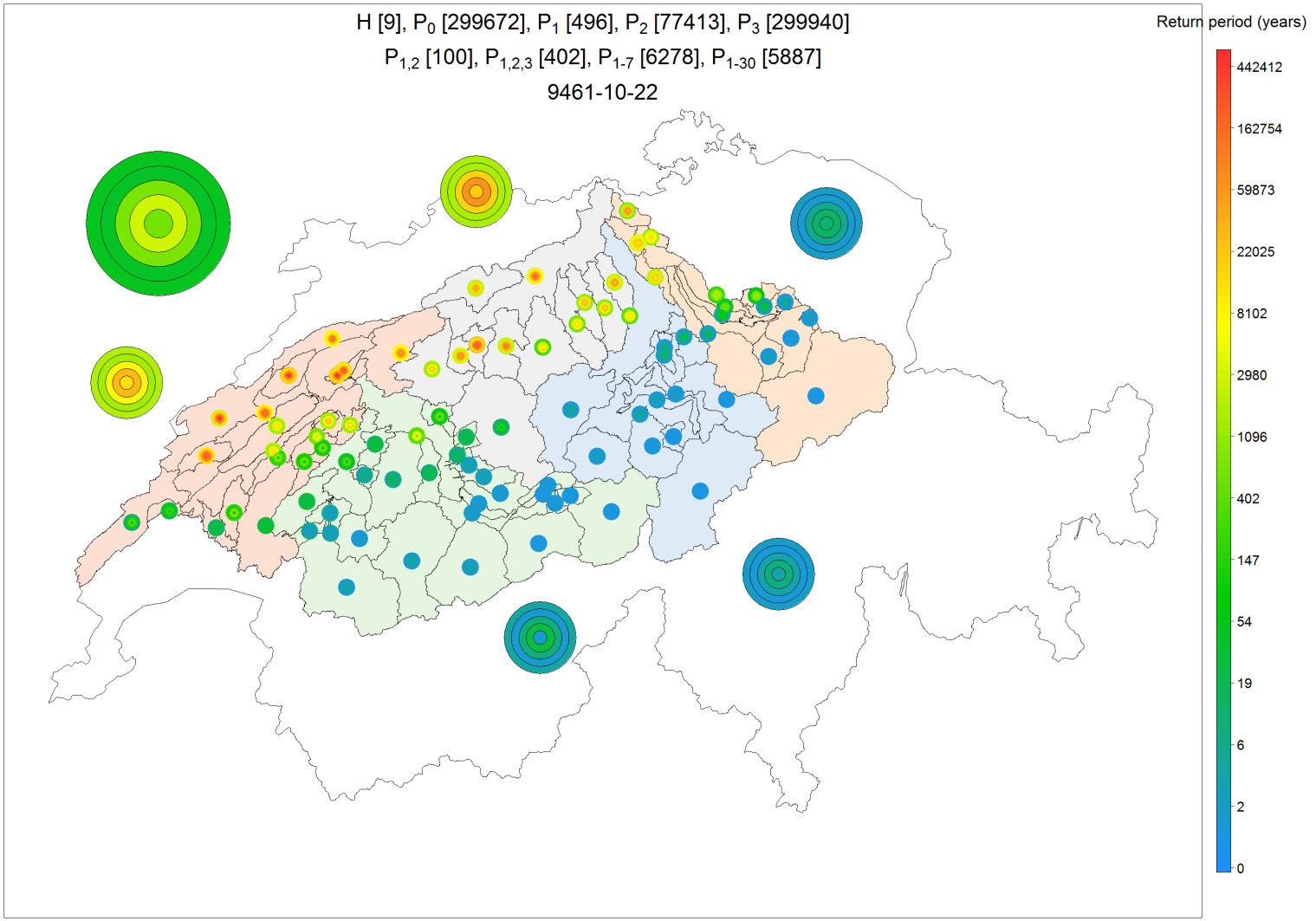


Figure A. 110 Severity map representing the severity at several spatial scales and for several accumulation periods for the 9thhydrological event

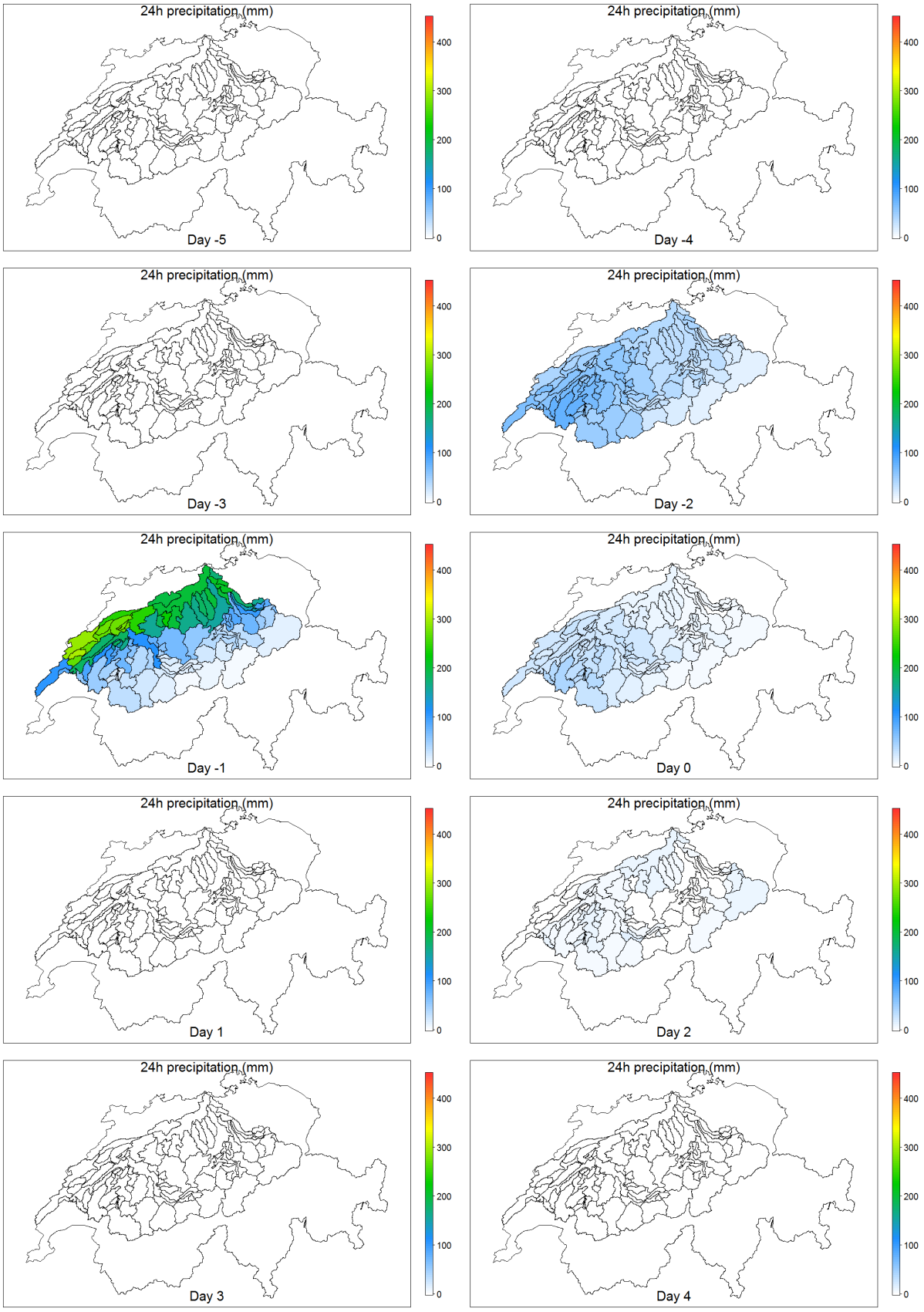


Figure A. 111 Space/time dynamics of precipitation during 10 days for the 9thhydrological event

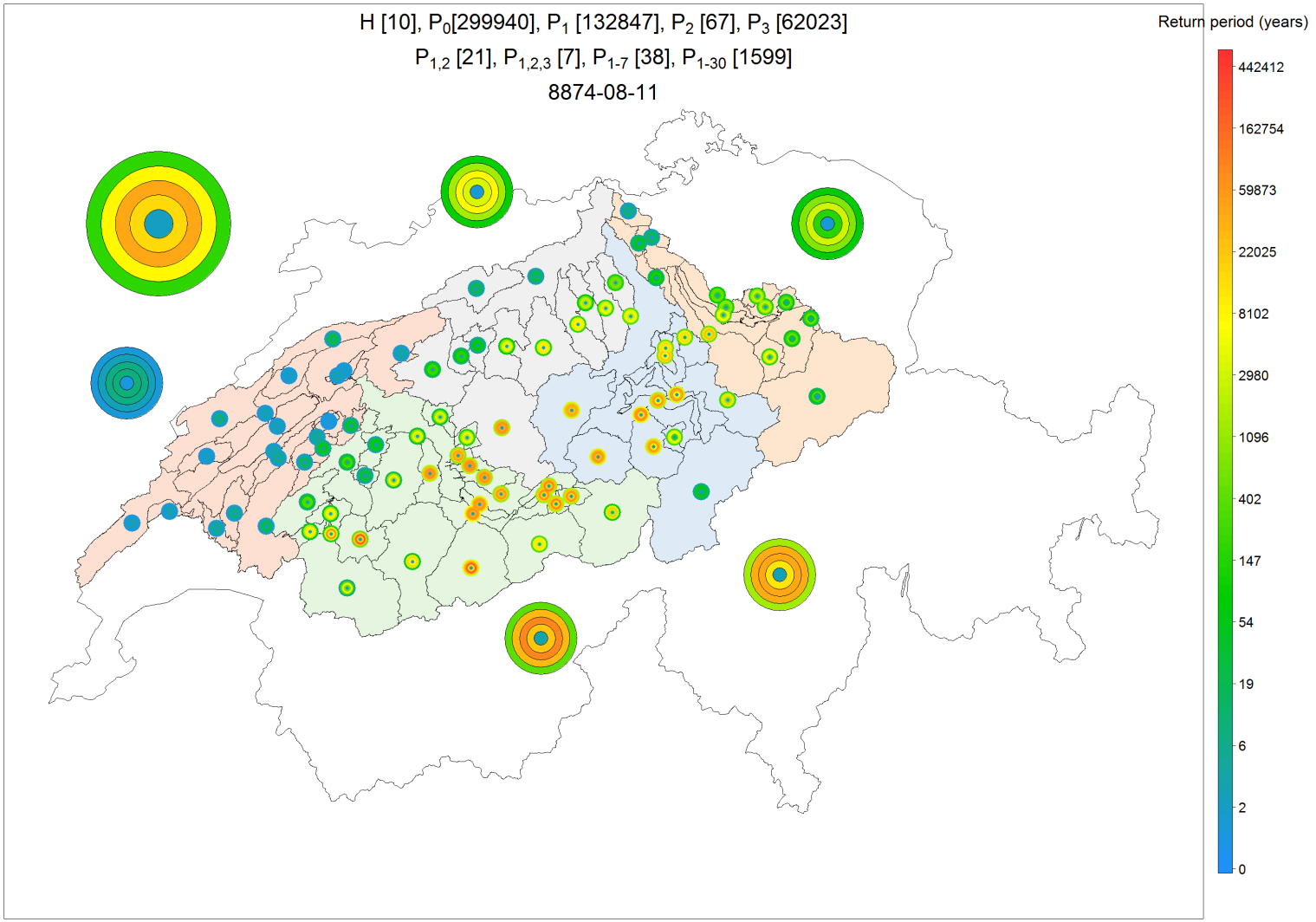


Figure A. 112 Severity map representing the severity at several spatial scales and for several accumulation periods for the 10th 1st hydrological event

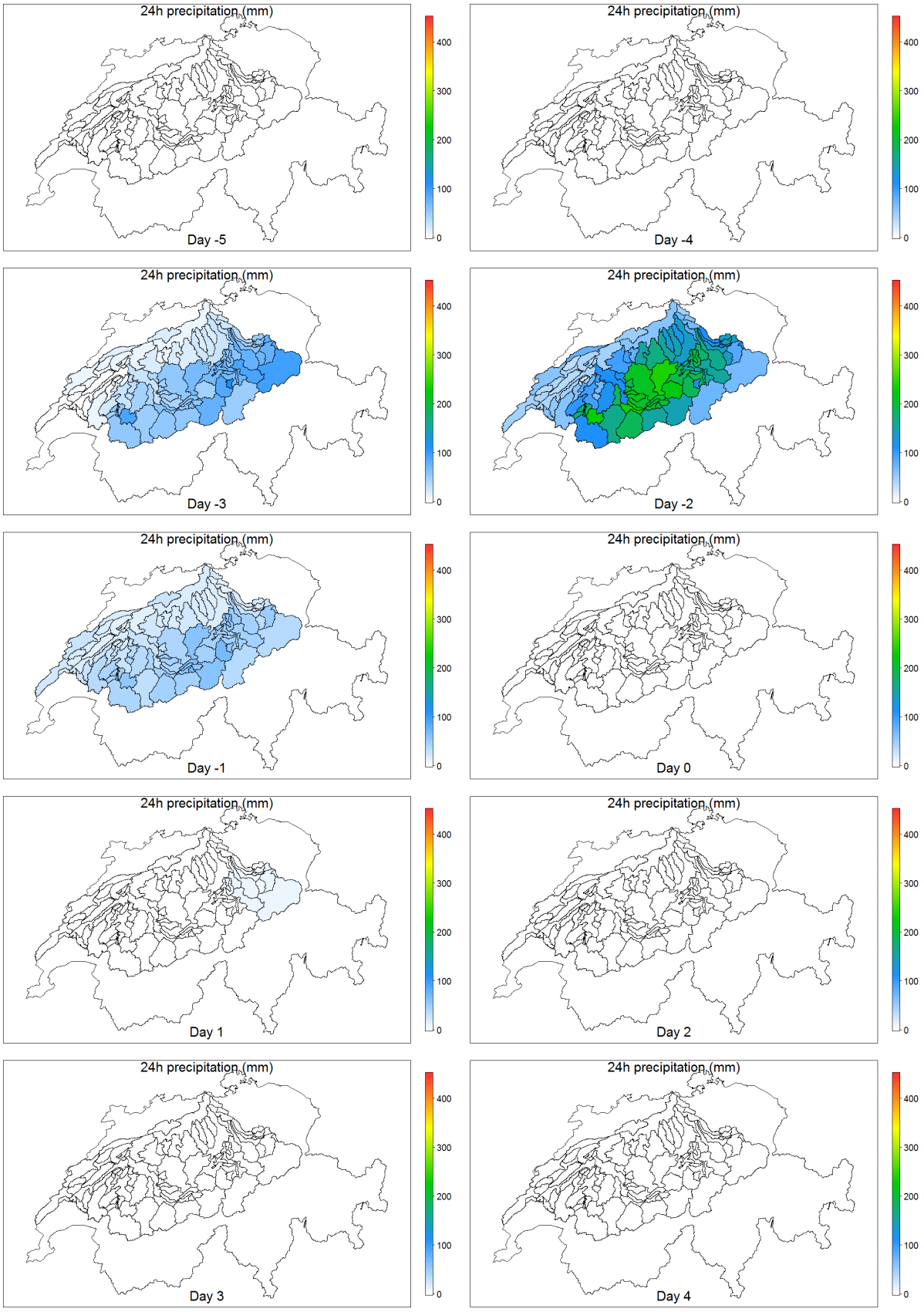


Figure A. 113 Space/time dynamics of precipitation during 10 days for the 10th hydrological event

10.6 PMP estimates and extremes generated with GWEX

JM. Fallot presented "PMP (Probable Maximum Precipitation) Maps for Switzerland" during the CRUEX workshop in March 2017. In this work, 3 and 24-hour PMP (Figure A. 114) are calculated by an Atmospheric model with horizontal resolution of 2 km for a number of different large-scale atmospheric configurations (wind speed / wind direction / atmosphere humidity…). For each pixel of the model, final PMP correspond to the highest value calculated for orographic and convergence PMP for the 12 wind directions. We have also to keep in mind that:

- According to Prof. Schleiss' assessment, they represent a return period somewhere in the range of 100'000–300'000 years > we have thus no clear estimate of their return period.
- These PMP estimates are uncertain as a result of the possible limitations of the atmospheric model and of its parametrization.
- Results of the GWEX and PMP approaches are not comparable.

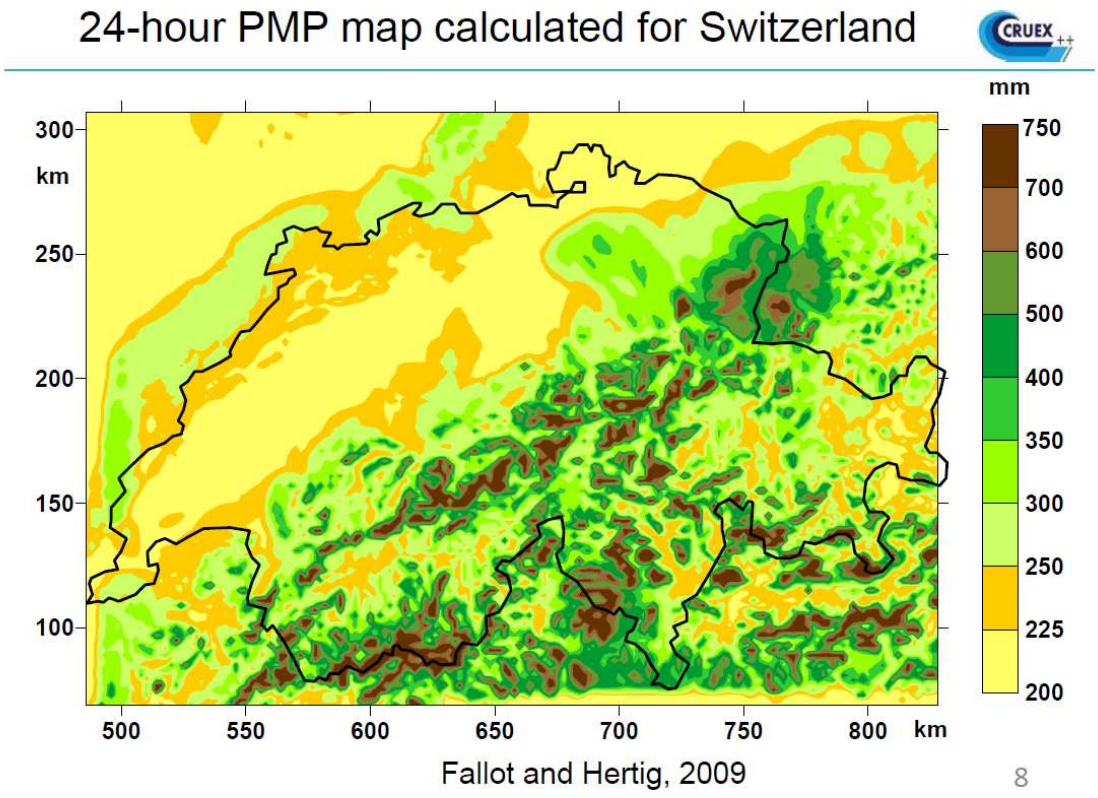


Figure A. 114 PMP estimate for each 2km pixel (Fallot, 2017).

Figure A. 115 and Figure A. 116 represent for the 105 stations the 300,000 year GWEX maxima and the GWEX maxima/PMP ratio, respectively. We observe that the spatial pattern of station maxima is similar to PMP with local differences. We find that several station maxima are above or below the PMP. In most cases the larger values seem to be not critical when PMP values in the neighbourhood are considered. However, the exceeding are very large for 4 stations (points surounded by red circles in figures). The impact of these very large values on basin precipitation are attenuated by a strong areal reduction factor effect (Figure A. 116).

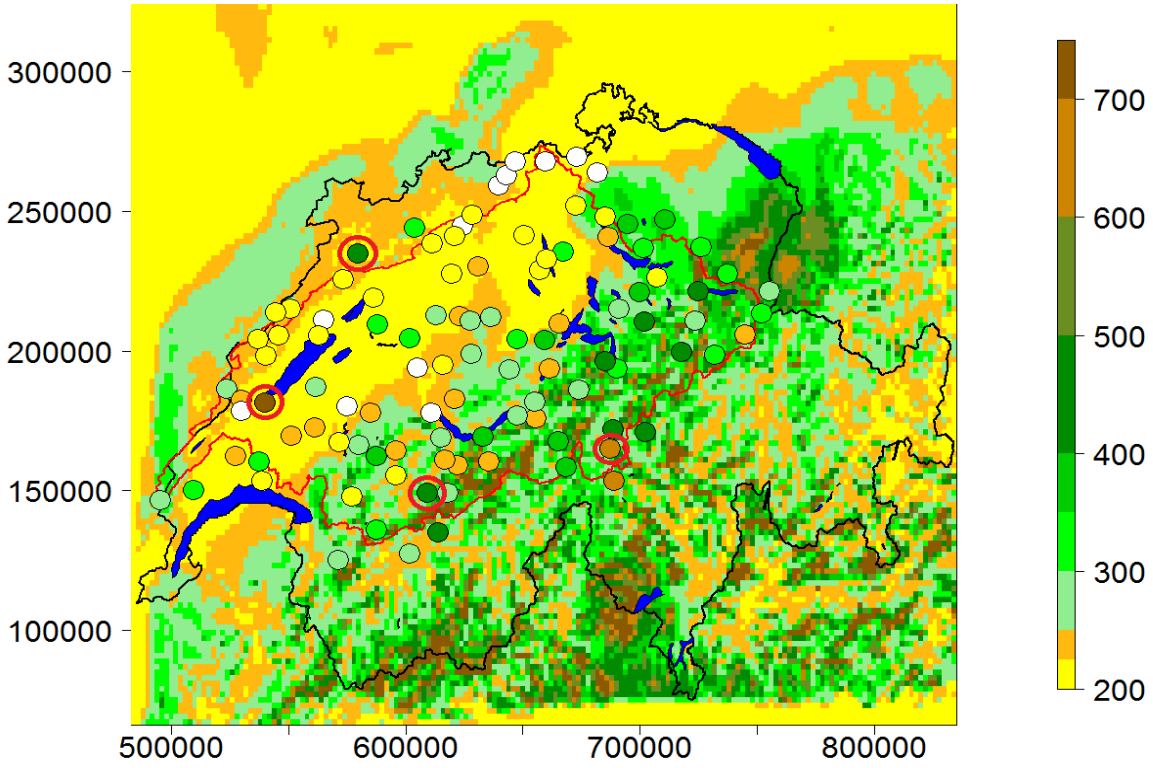


Figure A. 115 300'000 year GWEX maxima (colored circles) with PMP (background colors)

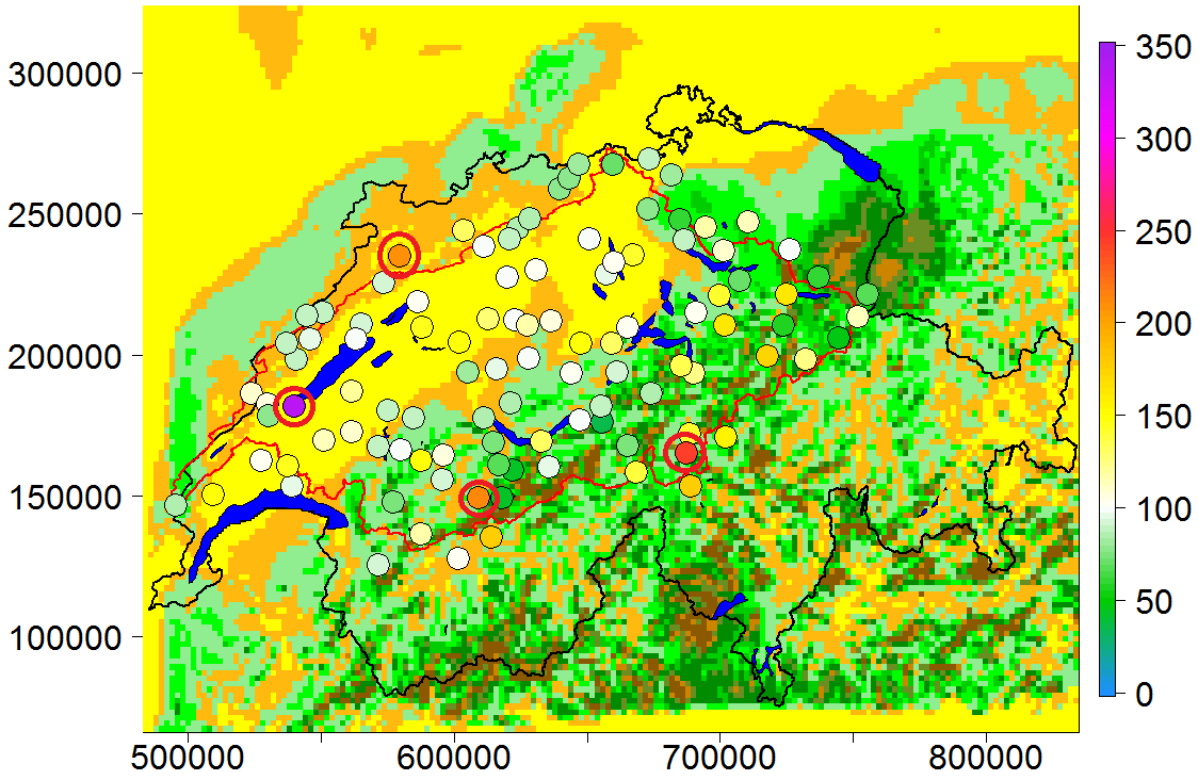


Figure A. 116 Ratio GWEX maxima/PMP in percentage.

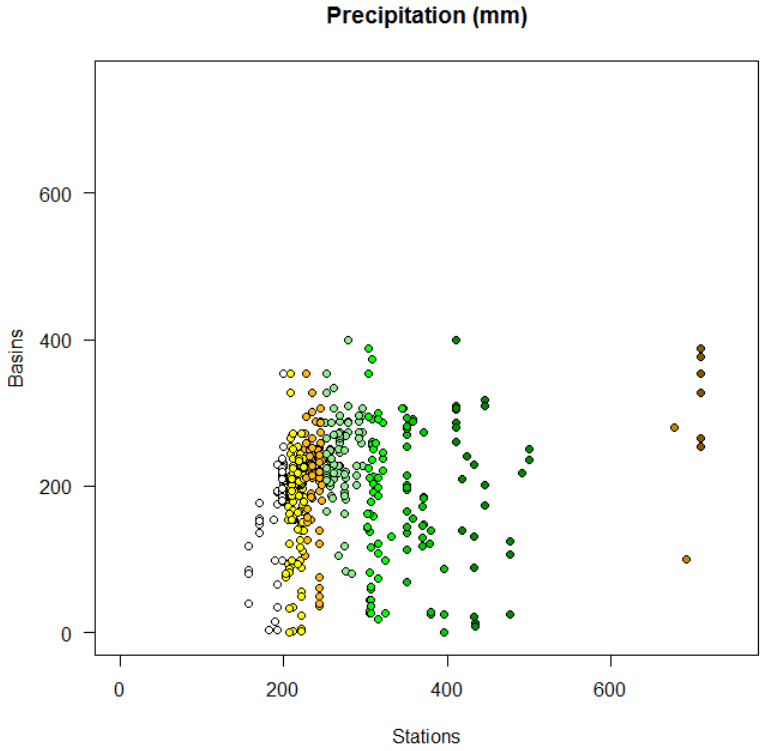


Figure A. 117 Basin maximum precipitation vs. station maximum precipitation.

At the catchment scale, the comparison between PMP and GWEX estimates is not possible anymore.

- The PMP maps map for Switzerland the PMP for each pixel as they could be observed based on the estimations from the model. It was mentioned in previous meetings that the maps should only be used for catchments with an area of up to 230 km² since a simultaneous PMP in a larger area were out of the question physically. PMP maps give thus no information on the mean areal precipitation amounts that could be obtained over some medium to large scale surface area. The pixel values of PMP that are shown on the PMP maps indeed often correspond to different large-scale atmospheric configurations (depending for instance on the orientation of hillslopes under consideration). Each of such configurations has no chance to give the PMP everywhere at the same time (or within a same meteorological event). This is clear in the presentation of JM Fallot: the 24-PMP in the southern part of the Swiss alps are obtained for SW to SE winds. In the northern part of the Swiss Alps, precipitation amounts obtained for those configurations are much lower whereas PMP are obtained for W, NW, NE or E winds (Figure A. 117Figure A. 116).
- Conversely, the EXAR weather generators were developed to produce weather scenarios that are statistically consistent in space. The spatial covariance structure of at site precipitation was identified from observations and used for the generation. The mean areal precipitation amounts obtained for any surface area / river basin / sub-catchment of the Aare are thus statistically realistic in the scenarios obtained with the weather generators. In the weather generators, the spatial covariance structure was even improved to account for different covariability characteristics for extremes.

We know from Figure A. 116 that the maximum precipitations simulated with GWEX for the 89 sub-catchments of the Aare are mostly significantly lower than the maximum precipitations obtained at each individual station.

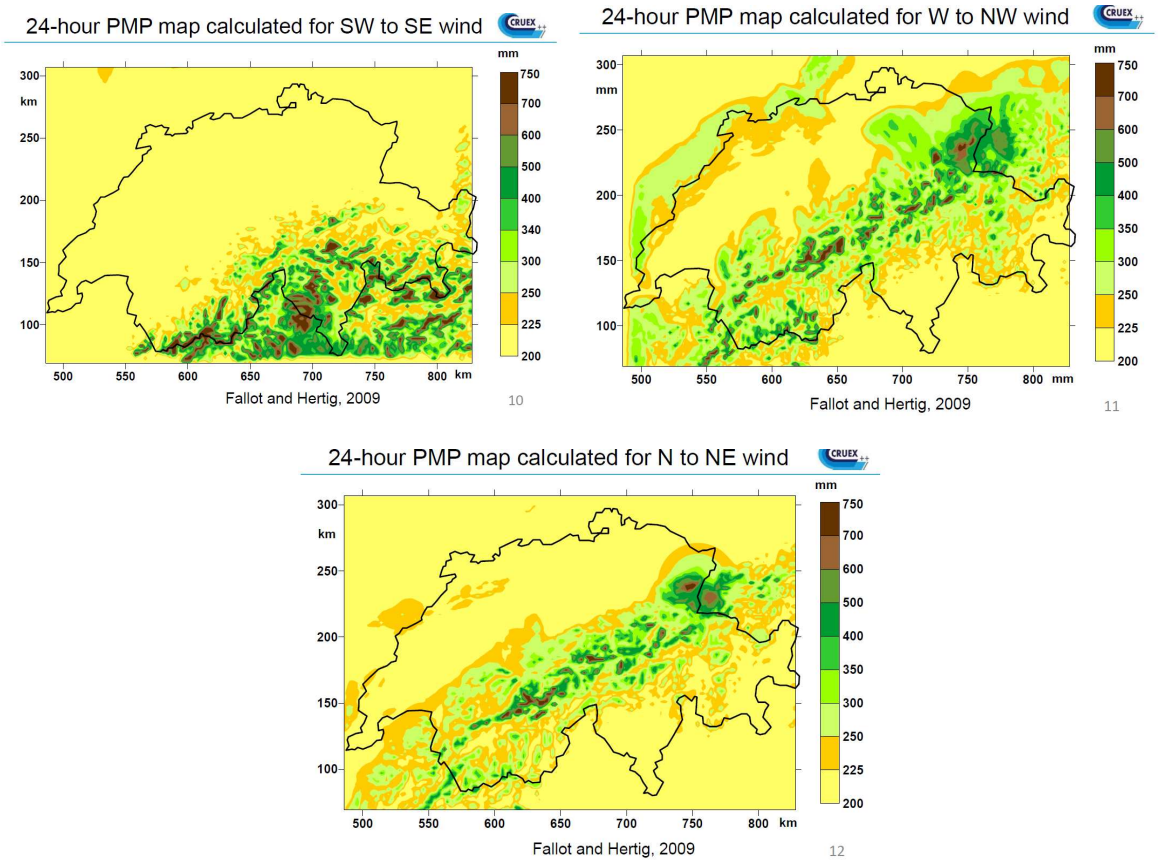


Figure A. 118 PMP obtained for different wind directions (Fallot, 2017).

References

Fallot, J.M. 2017. PMP (Probable Maximum Precipitation) Maps for Switzerland". Verification of flood safety for dams: Application of the CRUEX++ methodology for the estimation of extreme floods. CRUEX workshop, March 2017, Lausanne.

Zurich, 06.11.2020

University of Zurich

Dr. Maria Staudinger
Postdoc

PD Dr. Daniel Viviroli
Senior Scientist